SANDIA REPORT

SAND2003-8601
Unlimited Release
Printed October 2003

Mathematical Analysis of Deception

Nancy A. Durgin and Deanna K. Rogers

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2003-8601
Unlimited Release
Printed October 2003

Mathematical Analysis of Deception

Nancy Durgin
Department of Information Security
Sandia National Laboratories
P.O. Box 969
Livermore, CA 94551
nadurgi@sandia.gov

Deanna Koike Rogers
University of California at Davis
Davis, CA
koike@cs.ucdavis.edu

Abstract

This report describes the results of a three year research project about the use of deception in information protection. The work involved a collaboration between Sandia employees and students in the Center for Cyber Defenders (CCD) and at the University of California at Davis. This report includes a review of the history of deception, a discussion of some cognitive issues, an overview of previous work in deception, the results of experiments on the effects of deception on an attacker, and a mathematical model of error types associated with deception in computer systems.

This page intentionally left blank

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Executive Summary

In the last few years, deception has emerged as one of the key techniques for effective information protection in networks. A natural side effect of the use of this technology is the desire to understand the mathematical properties underlying its utility. Several informal notions have been introduced regarding this, for example:

- Deception increases the attacker's workload because they can't easily tell which of their attack attempts work and which fail.
- Deception allows defenders to track attacker attempts at entry and respond before attackers come across a vulnerability the defenders are susceptible to.
- Deception exhausts attacker resources.
- Deception increases the sophistication required for attack.
- Deception increases attacker uncertainty.

The goal of this project was to examine these claims and provide a more mathematical foundation for this aspect of deception as a tool for network defense. This collaborative research with the University of California at Davis and student interns from the Center for Cyber Defenders (CCD) benefits improved understanding that can be applied to the infrastructure of the DOE complex, since its results will prove beneficial to the wider national defense issues of computer networks.

The results of the research presented in this paper seem to support the common notions about deception that are outlined above. Deception techniques have the demonstrated ability to increase attacker workload and reduce attacker effectiveness. In addition, deception can decrease the defender effort required for detection and provide substantial increases in defender understanding of attacker capabilities and intent. Anecdotal evidence seems to indicate that even belief that deception technology is in use on a system can result in some benefit to the defender, because attackers don't trust their results and do more cross-checking, resulting in slower progress. Additional projects are already underway to deploy deception technologies to defend information systems at both the network and host level.

1.2 Overview

The bulk of this report comprises four papers that were written over the course of the three year project. These papers had originally been published electronically; they have been reformatted and typeset for this report, with the original authors and the approximate date of publishing left intact. In Chapter 2 we present a review of the history of deception, a discussion of some cognitive issues, and an overview of previous work done in this area. In Chapter 3 we present some experimental results on the effects of deception on Red Teaming exercises. In Chapter 4 we describe some experiments that demonstrate how it is possible to use deception to lead attackers through paths in an attack graph. In Chapter 5 we describe a mathematical model of error types associated with deception in computer systems. Finally, in Chapter 6 we present some concluding remarks.

Chapter 2

A Framework for Deception

By Fred Cohen, Dave Lambert, Charles Preston, Nina Berry, Corbin Stewart, and Eric Thomas[1]
July 13, 2001

- Fred Cohen: Sandia National Laboratories
- Dave Lambert: SPAWAR Systems Center
- Charles Preston: Information Integrity, University of New Haven
- Nina Berry: Sandia National Laboratories
- Corbin Stewart: Sandia National Laboratories (CCD)
- Eric Thomas: Sandia National Laboratories (CCD)

2.1 Abstract

This paper overviews issues in the use of deception for information protection. Its objective is to create a framework for deception and an understanding of what is necessary for turning that framework into a practical capability for carrying out defensive deceptions for information protection.

2.1.1 Overview of results

We have undertaken an extensive review of literature to understand previous efforts in this area and to compile a collection of information in areas that appear to be relevant to the subject at hand. It has become clear through this investigation that there is a great deal of additional detailed literature that should be reviewed in order to create a comprehensive collection. However, it appears that the necessary aspects of the subject have been covered and that additional collection will likely be comprised primarily of detailing in areas that are now known to be relevant.

We have developed a framework for creating and analyzing deceptions involving individual people, individual computers, one person acting with one computer, networks of people, networks of computers, and organizations consisting of people and their associated computers. This framework has been used to model select deceptions and, to a limited extent, to assist in the development of new deceptions. This framework is described in the body of this report with additional details provided in the appendixes.

Based on these results; (1) we are now able to understand and analyze deceptions with considerably more clarity than we could previously, (2) we have command of a far greater collection of

[1]This chapter is published online at `http://all.net/journal/deception/Framework/Framework.html`.

techniques available for use in defensive deception than was previously available and than others have published in the field, and (3) we now have a far clearer understanding of how and when to apply which sorts of techniques than was previously available. It appears that with additional effort over time we will be able to continue to develop greater and more comprehensive understanding of the subject and extend our understanding, capabilities, and techniques.

2.1.2 Further Work

It appears that a substantial follow-on effort is required in order to systematize the creation of defensive information protection deceptions. Such an effort would most likely require:

- The creation of a comprehensive collection of material on key subject areas related to deception. This has been started in this paper but there is clearly a great deal of effort left to be done.
- The creation of a database supporting the creation of analysis of defensive deceptions and a supporting software capability to allow that database to be used by experts in their creation and operation of deceptions.
- A team of experts working to create and maintain a capability for supporting deceptions and sets of supporting personnel used as required for the implementation of specific deceptions.

We strongly believe that this effort should continue over an extended period of time and with adequate funding, and that such effort will allow us to create and maintain a substantial lead over the threat types currently under investigation. The net effect will be an ongoing and increasing capability for the successful deception of increasingly skilled and hostile threats.

2.2 Introduction and Overview

According to the American Heritage Dictionary of the English Language (1981):

> "deception" is defined as "the act of deceit"
>
> "deceit" is defined as "deception".

Since long before 800 B.C. when Sun Tzu wrote "The Art of War" [Tzu83] deception has been key to success in warfare. Similarly, information protection as a field of study has been around for at least 4,000 years [Kah67] and has been used as a vital element in warfare. But despite the criticality of deception and information protection in warfare and the historical use of these techniques, in the transition toward an integrated digitized battlefield and the transition toward digitally controlled critical infrastructures, the use of deception in information protection has not been widely undertaken. Little study has apparently been undertaken to systematically explore the use of deception for protection of systems dependent on digital information. This paper, and the effort of which it is a part, seeks to change that situation.

In October of 1983 in explaining INFOWAR [Hub83], Robert E. Huber explains by first quoting from Sun Tzu:

> **Deception: The Key** The act of deception is an art supported by technology. When successful, it can have devastating impact on its intended victim. In Fact:
>
> "All warfare is based on deception. Hence, when able to attack, we must seem unable; when using our forces, we must seem inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near. Hold out baits to entice the enemy. Feign disorder, and crush him. If he is secure at all points, be prepared for him. If he is in superior strength, evade him. If your opponent is

> of choleric temper, seek to irritate him. Pretend to be weak, that he may grow arrogant. If he is taking his ease, give him no rest. If his forces are united, separate them. Attack him where he is unprepared, appear where you are not expected."[Tzu83]
>
> The ability to sense, monitor, and control own-force signatures is at the heart of planning and executing operational deception...
>
> The practitioner of deception utilizes the victim's intelligence sources, surveillance sensors and targeting assets as a principal means for conveying or transmitting a deceptive signature of desired impression. It is widely accepted that all deception takes place in the mind of the perceiver. Therefore it is *not* the act itself but the acceptance that counts!"

It seems to us at this time that there are only two ways of defeating an enemy:

1. One way is to have overwhelming force of some sort (i.e., an actual asymmetry that is, in time, fatal to the enemy). For example, you might be faster, smarter, better prepared, better supplied, better informed, first to strike, better positioned, and so forth.

2. The other way is to manipulate the enemy into reduced effectiveness (i.e., induced misperceptions that cause the enemy to misuse their capabilities). For example, the belief that you are stronger, closer, slower, better armed, in a different location, and so forth.

Having both an actual asymmetric advantage and effective deception increases your advantage. Having neither is usually fatal. Having more of one may help balance against having less of the other. Most military organizations seek to gain both advantages, but this is rarely achieved for long, because of the competitive nature of warfare.

2.2.1 Overview of This Paper

The purpose of this paper is to explore the nature of deception in the context of information technology defenses. While it can be reasonably asserted that all information systems are in many ways quite similar, there are differences between systems used in warfare and systems used in other applications, if only because the consequences of failure are extreme and the resources available to attackers are so high. For this reason, military situations tend to be the most complex and risky for information protection and thus lead to a context requiring extremes in protective measures. When combined with the rich history of deception in warfare, this context provides fertile ground for exploring the underlying issues.

We begin by exploring the history of deception and deception techniques. Next we explore the nature of deception and provide a set of dimensions of the deception problem that are common to deceptions of the targets of interest. We then explore a model for deception of humans, a model for deception of computers, and a set of models of deceptions of systems of people and computers. Finally, we consider how we might design and analyze deceptions, discuss the need for experiments in this arena, summarize, draw conclusions, and describe further work.

2.3 A Short History of Deception

2.3.1 Deception in Nature

While Sun Tzu is the first known publication depicting deception in warfare as an art, long before Sun Tzu there were tribal rituals of war that were intended in much the same way. The beating of chests [Kee93] is a classic example that we still see today, although in a slightly different form. Many animals display their apparent fitness to others as part of the mating ritual of for territorial assertions [MT86]. Mitchell and Thompson look at human and nonhuman deception and provide interesting perspectives from many astute authors on many aspects of this subject. We see much the same behavior in today's international politics. Who could forget Kruschev banging his shoe on the table at the UN and declaring "We will bury you!" Of course it's not only the losers that 'beat their chests', but it is a more stark example if presented that way. Every nation declares its greatness, both to its own people and to the world at large. We may call it pride, but at some point it becomes bragging, and in conflict situations, it becomes a display. Like the ancient tribesmen, the goal is, in some sense, to avoid a fight. The hope is that, by making the competitor think that it is not worth taking us on, we will not have to waste our energy or our blood in fighting when we could be spending it in other ways. Similar noise-making tactics also work to keep animals from approaching an encampment. The ultimate expression of this is in the area of nuclear deterrence [Wil68].

Animals also have genetic characteristics that have been categorized as deceptions. For example, certain animals are able to change colors to match the background or, as in the case of certain types of octopi, the ability to mimic other creatures. These are commonly lumped together, but in fact they are very different. The moth that looks like a flower may be able to 'hide' from birds but this is not an intentional act of deception. Survival of the fittest simply resulted in the death of most of the moths that could be detected by birds. The ones that happened to carry a genetic trait that made them look like a particular flower happened to get eaten less frequently. This is not a deception, it is a trait that survives. The same is true of the Orca whale which has colors that act as a dazzlement to break up its shape.

On the other hand, anyone who has seen an octopus change coloring and shape to appear as if it were a rock when a natural enemy comes by and then change again to mimic a food source while lying in wait for a food source could not honestly claim that this was an unconscious effort. This form of concealment (in the case of looking like a rock or foodstuff) or simulation (in the case of looking like an inedible or hostile creature) is highly selective, driven by circumstance, and most certainly driven by a thinking mind of some sort. It is a deception that uses a genetically endowed physical capability in an intentional and creative manner. It is more similar to a person putting on a disguise than it is to a moth's appearance.

2.3.2 Historical Military Deception

The history of deception is a rich one. In addition to the many books on military history that speak to it, it is a basic element of strategy and tactics that has been taught since the time of Sun Tzu. But in many ways, it is like the history of biology before genetics. It consists mainly of a collection of examples loosely categorized into things that appear similar at the surface. Hiding behind a tree is thought to be similar to hiding in a crowd of people, so both are called concealment. On the surface they appear to be the same, but if we look at the mechanisms underlying them, they are quite different.

> Historically, military deception has proven to be of considerable value in the attainment of national security objectives, and a fundamental consideration in the development and implementation of military strategy and tactics. Deception has been used to enhance, exaggerate, minimize, or distort capabilities and intentions; to mask deficiencies; and to otherwise cause desired appreciations where conventional military activities and security measures were unable to achieve the desired result. The development of a deception organization and the exploitation of deception opportunities are considered to be vital to national security. To develop deception capabilities, including procedures and techniques for deception staff components, it is essential that deception receive continuous command emphasis in military exercises, command post exercises, and in training operations." – JCS Memorandum of Policy (MOP) 116 [Arm98]

MOP 116 also points out that the most effective deceptions exploit beliefs of the target of the deception and, in particular, decision points in the enemy commander's operations plan. By altering the enemy commander's perception of the situation at key decision points, deception may turn entire campaigns.

There are many excellent collections of information on deceptions in war. One of the most comprehensive overviews comes from Whaley [Wha69], which includes details of 67 military deception operations between 1914 and 1968. The appendix to Whaley is 628 pages long and the summary charts (in appendix B) are another 50 pages. Another 30 years have passed since this time, which means that it is likely that another 200 pages covering 20 or so deceptions should be added to update this study. Dunnigan and Nofi [DN95] review the history of deception in warfare with an eye toward categorizing its use. They identify the different modes of deception as concealment, camouflage, false and planted information, ruses, displays, demonstrations, feints, lies, and insight.

Dewar [Dew89] reviews the history of deception in warfare and, in only 12 pages, gives one of the most cogent high-level descriptions of the basis, means, and methods of deception. In these 12 pages, he outlines (1) the weaknesses of the human mind (preconceptions, tendency to think we are right, coping with confusion by leaping to conclusions, information overload and resulting filtering, the tendency to notice exceptions and ignore commonplace things, and the tendency to be lulled by regularity), (2) the object of deception (getting the enemy to do or not do what you wish), (3) means of deception (affecting observables to a level of fidelity appropriate to the need, providing consistency, meeting enemy expectations, and not making it too easy), (4) principles of deception (careful centralized control and coordination, proper preparation and planning, plausibility, the use of multiple sources and modes, timing, and operations security), and (5) techniques of deception (encouraging belief in the most likely when a less likely is to be used, luring the enemy with an ideal opportunity , the repetitive process and its lulling effect, the double bluff which involves revealing the truth when it is expected to be a deception, the piece of bad luck which the enemy believes they are taking advantage of, the substitution of a real item for a detected deception item, and disguising as the enemy). He also (6) categorizes deceptions in terms of senses and (7) relates 'security' (in which you try to keep the enemy from finding anything out) to deception (in which you try to get the enemy to find out the thing you want them to find). Dewar includes pictures and examples in these 12 pages to boot.

In 1987, Knowledge Systems Corporation [Kno87] created a useful set of diagrams for planning tactical deceptions. Among their results, they indicate that the assessment and planning process is manual, lacks automated applications programs, and lacks timely data required for combat support. This situation does not appear to have changed. They propose a planning process consisting of (1) reviewing force objectives, (2) evaluating your own and enemy capabilities and other situational factors, (3) developing a concept of operations and set of actions, (4) allocating resources, (5) coordinating and deconflicting the plan relative to other plans, (6) doing a risk and feasibility assessment, (7) reviewing adherence to force objectives, and (8) finalizing the plan. They detail steps to accomplish each of these tasks in useful process diagrams and provide forms for doing a more systematic analysis of deceptions than was previously available. Such a planning mechanism does not appear to exist today for deception in information operations.

These authors share one thing in common. They all carry out an exercise in building categories. Just as the long standing effort of biology to build up genus and species based on bodily traits (phenotypes), eventually fell to a mechanistic understanding of genetics as the underlying cause, the scientific study of deception will eventually yield a deeper understanding that will make the mechanisms clear and allow us to understand and create deceptions as an engineering discipline. That is not to say that we will necessarily achieve that goal in this short examination of the subject, but rather that in-depth study will ultimately yield such results.

There have been a few attempts in this direction. A RAND study included a 'straw man' graphic [Gri78](H7076) that showed deception as being broken down into "Simulation" and "Dissimulation Camouflage".

> Whaley first distinguishes two categories of deception (which he defines as one's intentional distortion of another's perceived reality): 1) dissimulation (hiding the real) and

2) simulation (showing the false). Under dissimulation he includes: a) masking (hiding the real by making it invisible), b) repackaging (hiding the real by disguising), and c) dazzling (hiding the real by confusion). Under simulation he includes: a) mimicking (showing the false through imitation), b) inventing (showing the false by displaying a different reality), and c) decoying (showing the false by diverting attention). Since Whaley argues that "everything that exists can to some extent be both simulated and dissimulated, " whatever the actual empirical frequencies, at least in principle hoaxing should be possible for any substantive area."[Ste93, page 293]

The same slide reflects on Dewar's view [Dew89] that security attempts to deny access and counterintelligence attempts while deception seeks to exploit intelligence. Unfortunately, the RAND depiction is not as cogent as Dewar in breaking down the 'subcategories' of simulation. The RAND slides do cover the notions of observables being "known and unknown", "controllable and uncontrollable", and "enemy observable and enemy non-observable". This characterization of part of the space is useful from a mechanistic viewpoint and a decision tree created from these parameters can be of some use. Interestingly, RAND also points out the relationship of selling, acting, magic, psychology, game theory, military operations, probability and statistics, logic, information and communications theories, and intelligence to deception. It indicates issues of observables, cultural bias, knowledge of enemy capabilities, analytical methods, and thought processes. It uses a reasonable model of human behavior, lists some well known deception techniques, and looks at some of the mathematics of perception management and reflexive control.

2.3.3 Cognitive Deception Background

Many authors have examined facets of deception from both an experiencial and cognitive perspective.

Chuck Whitlock has built a large part of his career on identifying and demonstrating these sorts of deceptions [Whi97]. His book includes detailed descriptions and examples of scores of common street deceptions. Fay Faron points out that most such confidence efforts are carried as as specific 'plays' and details the anatomy of a 'con' [Far98]. She provides 7 ingredients for a con (too good to be true, nothing to lose, out of their element, limited time offer, references, pack mentality, and no consequence to actions). The anatomy of the confidence game is said to involve (1) a motivation (e.g., greed), (2) the come-on (e.g., opportunity to get rich), (3) the shill (e.g., a supposedly independent third party), (4) the swap (e.g., take the victim's money while making them think they have it), (5) the stress (e.g., time pressure), and (6) the block (e.g., a reason the victim will not report the crime). She even includes a 10-step play that makes up the big con.

Bob Fellows [Fel00] takes a detailed approach to how 'magic' and similar techniques exploit human fallibility and cognitive limits to deceive people. According to Fellows (p14) the following characteristics improve the changes of being fooled: (1) under stress, (2) naivety, (3) in life transitions, (4) unfulfilled desire for spiritual meaning, (5) tend toward dependency, (6) attracted to trance-like states of mind, (7) unassertive, (8) unaware of how groups can manipulate people, (9) gullible, (10) have had a recent traumatic experience, (11) want simple answers to complex questions, (12) unaware of how the mind and body affect each other, (13) idealistic, (14) lack critical thinking skills, (15) disillusioned with the world or their culture, and (16) lack knowledge of deception methods. Fellows also identifies a set of methods used to manipulate people.

Thomas Gilovich [Gil91] provides in-depth analysis of human reasoning fallibility by presenting evidence from psychological studies that demonstrate a number of human reasoning mechanisms resulting in erroneous conclusions. This includes the general notions that people (erroneously) (1) believe that effects should resemble their causes, (2) misperceive random events, (3) misinterpret incomplete or unrepresentative data, (4) form biased evaluations of ambiguous and inconsistent data, (5) have motivational determinants of belief, (6) bias second hand information, and (7) have exaggerated impressions of social support. Substantial further detailing shows specific common syndromes and circumstances associated with them.

Charles K. West [Wes81] describes the steps in psychological and social distortion of information and provides detailed support for cognitive limits leading to deception. Distortion comes from the fact of an unlimited number of problems and events in reality, while human sensation can only sense certain types of events in limited ways: (1) A person can only perceive a limited number of those events at any moment, (2) A person's knowledge and emotions partially determine which of the events are noted and interpretations are made in terms of knowledge and emotion (3) Intentional bias occurs as a person consciously selects what will be communicated to others, and (4) the receiver of information provided by others will have the same set of interpretations and sensory limitations.

Al Seckel [Sec00] provides about 100 excellent examples of various optical illusions, many of which work regardless of the knowledge of the observer, and some of which are defeated after the observer sees them only once. Donald D. Hoffman [Hof98] expands this into a detailed examination of visual intelligence and how the brain processes visual information. It is particularly noteworthy that the visual cortex consumes a great deal of the total human brain space and that it has a great deal of effect on cognition. Some of the 'rules' that Hoffman describes with regard to how the visual cortex interprets information include: (1) Always interpret a straight line in an image as a straight line in 3D, (2) If the tips of two lines coincide in an image interpret them as coinciding in 3D, (3) Always interpret co-linear lines in an image as co-linear in 3D, (4) Interpret elements near each other in an image as near each other in 3D, (5) Always interpret a curve that is smooth in an image as smooth in 3D, (6) Where possible, interpret a curve in an image as the rim of a surface in 3D, (7) Where possible, interpret a T-junction in an image as a point where the full rim conceals itself; the cap conceals the stem, (8) Interpret each convex point on a bound as a convex point on a rim, (9) Interpret each concave point on a bound as a concave point on a saddle point, (10) Construct surfaces in 3D that are as smooth as possible, (11) Construct subjective figures that occlude only if there are convex cusps, (12) If two visual structures have a non-accidental relation, group them and assign them to a common origin, (13) If three or more curves intersect at a common point in an image, interpret them as intersecting at a common point in space, (14) Divide shapes into parts along concave creases, (15) Divide shapes into parts at negative minima, along lines of curvature, of the principal curvatures, (16) Divide silhouettes into parts at concave cusps and negative minima of curvature, (17) The salience of a cusp boundary increases with increasing sharpness of the angle at the cusp, (18) The salience of a smooth boundary increases with the magnitude of (normalized) curvature at the boundary, (19) Choose figure and ground so that figure has the more salient part boundaries, (20) Choose figure and ground so that figure has the more salient parts, (21) Interpret gradual changes in hue, saturation, and brightness in an image as changes in illumination, (22) Interpret abrupt changes in hue, saturation, and brightness in an image as changes in surfaces, (23) Construct as few light sources as possible, (24) Put light sources overhead, (25) Filters don't invert lightness, (26) Filters decrease lightness differences, (27) Choose the fair pick that's most stable, (28) Interpret the highest luminance in the visual field as white, flourent, or self-luminous, (29) Create the simplest possible motions, (30) When making motion, construct as few objects as possible, and conserve them as much as possible, (31) Construct motion to be as uniform over space as possible, (32) Construct the smoothest velocity field, (33) If possible, and if other rules permit, interpret image motions as projections of rigid motions in three dimensions, (34) If possible, and if other rules permit, interpret image motions as projections of 3D motions that are rigid and planar, (35) Light sources move slowly.

It appears that the rules of visual intelligence are closely related to the results of other cognitive studies. It may not be a coincidence that the thought processes that occupy the same part of the brain as visual processing have similar susceptibilities to errors and that these follow the pattern of the assumption that small changes in observation point should not change the interpretation of the image. It is surprising when such a change reveals a different interpretation, and the brain appears to be designed to minimize such surprises while acting at great speed in its interpretation mechanisms. For example, rule 2 (If the tips of two lines coincide in an image interpret them as coinciding in 3D) is very nearly always true in the physical world because coincidence of line ends that are not in fact coincident in 3 dimensions requires that you be viewing the situation at precisely the right angle with respect to the two lines. Another way of putting this is that there is a single line in space that connects the two points so as to make them appear to be coincident if they are not in fact conincident. If the observer is not on that single line, the points will not appear coincident. Since people usually have two eyes and they cannot align on the same line in space with respect to anything they can observe, there is no real 3 dimensional situation in which this coincidence can actually occur, it can only be simulated by 3 dimensional objects that are far enough away to appear to be on the same line with respect to both eyes, and there are no commonly occuring natural phenomena that pose anything of immediate visual import or consequence at thast distance. Designing visual stimuli that violate these principles will confuse most human observers and effective visual simulations should take these rules into account.

Deutsch [Deu95] provides a series of demonstrations of interpretation and misinterpretation of audio information. This includes: (1) the creation of words and phrases out of random sounds, (2) the susceptibility of interpretation to predisposition, (3) misinterpretation of sound based on relative pitch of pairs of tones, (4) misinterpretation of direction of sound source based on switching speakers, (5) creation of different words out of random sounds based on rapid changes in source direction, and (6) the change of word creation over time based on repeated identical audio stimulus.

First Karrass [Kar70] then Cialdini [Cia01] have provided excellent summaries of negotiation strategies and the use of influence to gain advantage. Both also explain how to defend against influence tactics. Karrass was one of the early experimenters in how people interact in negotiations and identified (1) credibility of the presenter, (2) message content and appeal, (3) situation setting and rewards, and (4) media choice for messages as critical components of persuasion. He also identifies goals, needs, and perceptions as three dimensions of persuasion and lists scores of tactics categorized into types including (1) timing, (2) inspection, (3) authority, (4) association, (5) amount, (6) brotherhood, and (7) detour. Karrass also provides a list of negotiating techniques including: (1) agendas, (2) questions, (3) statements, (4) concessions, (5) commitments, (6) moves, (7) threats, (8) promises, (9) recess, (10) delays, (11) deadlock, (12) focal points, (13) standards, (14) secrecy measures, (15) nonverbal communications, (16) media choices, (17) listening, (18) caucus, (19) formal and informal memorandum, (20) informal discussions, (21) trial balloons and leaks, (22) hostility releivers, (23) temporary intermediaries, (24) location of negotiation, and (25) technique of time.

Cialdini [Cia01] provides a simple structure for influence and asserts that much of the effect of influence techniques is built-in and occurs below the conscious level for most people. His structure consists of reciprocation, contrast, authority, commitment and consistency, automaticity, social proof, liking, and scarcity. He cites a substantial series of psychological experiments that demonstrate quite clearly how people react to situations without a high level of reasoning and explains how this is both critical to being effective decision makers and results in exploitation through the use of compliance tactics. While Cialdini backs up this information with numerous studies, his work is largely based on and largely cites western culture. Some of these elements are apparently culturally driven and care must be taken to assure that they are used in context.

Robertson and Powers [RP90] have worked out a more detailed low-level theoretical model of cognition based on "Perceptual Control Theory" (PCT), but extensions to higher levels of cognition have been highly speculative to date. They define a set of levels of cognition in terms of their order in the control system, but beyond the lowest few levels they have inadequate basis for asserting that these are orders of complexity in the classic control theoretical sense. The levels they include are intensity, sensation, configuration, transition / motion, events, relationships, categories, sequences / routines, programs / branching pathways / logic, and system concept.

David Lambert [Lam87] provides an extensive collection of examples of deceptions and deceptive techniques mapped into a cognitive model intended for modeling deception in military situations. These are categorized into cognitive levels in Lambert's cognitive model. The levels include sense, perceive feature, perceive form, associate, define problem / observe, define problem solving status (hypothesize), determine solution options, initiate actions / responses, direct, implement form, implement feature, and drive affectors. There are feedback and cross circuiting mechanisms to allow for reflexes, conditioned behavior, intuition, the driving of perception to higher and lower levels, and models of short and long term memory.

Charles Handy [Han93] discusses organizational structures and behaviors and the roles of power and influence within organizations. The National Research Council [NRC98] discusses models of human and organizational behavior and how automation has been applied in this area. Handy models organizations in terms of their structure and the effects of power and influence. Influence mechanisms are described in terms of who can apply them in what circumstances. Power is derived from physicality, resources, position (which yields information, access, and right to organize), expertise, personal charisma, and emotion. These result in influence through overt (force, exchange, rules and procedures, and persuasion), covert (ecology and magnetism), and bridging (threat of force) influences. Depending on the organizational structure and the relative positions of the participants, different aspects of power come into play and different techniques can be applied. The NRC report includes scores of examples of modeling techniques and details of simulation implementations based on those models and their applicability to current and future needs. Greene [Gre98] describes the 48 laws of power and, along the way, demonstrates 48 methods that exert compliance forces in an organization. These can be traced to cognitive influences and mapped out using models like Lambert's, Cialdini's, and the one we are considering for this effort.

Closely related to the subject of deception is the work done by the CIA on the MKULTRA project [MKU]. In June 1977, a set of MKULTRA documents were discovered, which had escaped destruction by the CIA. The Senate Select Committee on Intelligence held a hearing on August 3, 1977 to question CIA officials on the newly-discovered documents. The net effect of efforts to reveal information about this project was a set of released information on the use of sonic waves, electroshock, and other similar methods for altering peoples' perception. Included in this are such items as sound frequencies that make people fearful, sleepy, uncomfortable, and sexually aroused; results on hypnosis, truth drugs, psychic powers, and subliminal persuasion; LSD-related and other drug experiments on unwitting subjects; the CIA's "manual on trickery"; and so forth. One 1955 MKULTRA document gives an indication of the size and range of the effort; the memo refers to the study of an assortment of mind-altering substances which would: (1) "promote illogical thinking and impulsiveness to the point where the recipient would be discredited in public", (2) "increase the efficiency of mentation and perception", (3) "prevent or counteract the intoxicating effect of alcohol" (4) "promote the intoxicating effect of alcohol", (5) "produce the signs and symptoms of recognized diseases in a reversible way so that they may be used for malingering, etc." (6) "render the indication of hypnosis easier or otherwise enhance its usefulness" (7) "enhance the ability of individuals to withstand privation, torture and coercion during interrogation and so-called 'brainwashing', (8) "produce amnesia for events preceding and during their use", (9) "produce shock and confusion over extended periods of time and capable of surreptitious use", (10) "produce physical disablement such as paralysis of the legs, acute anemia, etc.", (11) "produce 'pure' euphoria with no subsequent let-down", (12) "alter personality structure in such a way that the tendency of the recipient to become dependent upon another person is enhanced", (13) "cause mental confusion of such a type that the individual under its influence will find it difficult to maintain a fabrication under questioning", (14) "lower the ambition and general working efficiency of men when administered in undetectable amounts", and (15) "promote weakness or distortion of the eyesight or hearing faculties, preferably without permanent effects".

A good summary of some of the pre-1990 results on psychological aspects of self-deception is provided in Heuer's CIA book on the psychology of intelligence analysis [Heu99]. Heuer goes one step further in trying to start assessing ways to counter deception, and concludes that intelligence analysts can make improvements in their presentation and analysis process. Several other papers on deception detection have been written and substantially summarized in Vrij's book on the subject [Vri00].

2.3.4 Computer Deception Background

In the early 1990s, the use of deception in defense of information systems came to the forefront with a paper about a deception 'Jail' created in 1991 by AT&T researchers in real-time to track an attacker and observe their actions [Che91]. An approach to using deceptions for defense by customizing every system to defeat automated attacks was published in 1992 [Coh92], while in 1996, descriptions of Internet Lightning Rods were given [Coh96b] and an example of the use of perception management to counter perception management in the information infrastructure was given [Coh96a]. More thorough coverage of this history was covered in a 1999 paper on the subject [Coh99b]. Since that time, deception has increasingly been explored as a key technology area for innovation in information protection. Examples of deception-based information system defenses include concealed services, encryption, feeding false information, hard-to-guess passwords, isolated sub-file-system areas, low building profile, noise injection, path diversity, perception management, rerouting attacks, retaining confidentiality of security status information, spread spectrum, and traps. In addition, it appears that criminals seek certainty in their attacks on computer systems and increased uncertainty caused by deceptions may have a deterrent effect [Coh98c].

The public release of DTK Deception ToolKit [Coh98a] led to a series of follow-on studies, technologies, and increasing adoption of technical deceptions for defense of information systems. This includes the creation of a small but growing industry with several commercial deception products, the HoneyNet project, the RIDLR project at Naval Post Graduate School, NSA-sponsored studies at RAND, the D-Wall technology [Coh00a, Coh99a], and a number of studies and developments now underway.

- **Commercial Deception Products:** The dominant commercial deception products today are DTK and Recourse Technologies. While the market is very new it is developing at a substantial rate and new results from deception projects are leading to an increased appreciation of the utility of deceptions for defense and a resulting increased market presence.

- **The HoneyNet Project:** The HoneyNet project is dedicated to learning and to the tools, tactics, and motives of the blackhat community and sharing the lessons learned. The primary tool used to gather this information is the Honeynet; a network of production systems designed to be compromised. This project has been joined by a substantial number of individual researchers and has had substantial success at providing information on widespread attacks, including the detection of large-scale denial of service worms prior to the use of the 'zombies' for attack. At least one Masters thesis is currently under way based on these results.

- **The RIDLR:** The RIDLR is a project launched from Naval Post Graduate School designed to test out the value of deception for detecting and defending against attacks on military information systems. RIDLR has been tested on several occasions at the Naval Post Graduate School and members of that team have participated in this project to some extent. There is an ongoing information exchange with that team as part of this project's effort.

- **RAND Studies:**

  In 1999, RAND completed an initial survey of deceptions in an attempt to understand the issues underlying deceptions for information protection [GRA99]. This effort included a historical study of issues, limited tool development, and limited testing with reasonably skilled attackers. The objective was to scratch the surface of possibilities and assess the value of further explorations. It predominantly explored intelligence related efforts against systems and methods for concealment of content and creation of large volumes of false content. It sought to understand the space of friendly defensive deceptions and gain a handle on what was likely to be effective in the future.

  > This report indicates challenges for the defensive environment including: (1) adversary initiative, (2) response to demonstrated adversary capabilities or established

friendly shortcomings, (3) many potential attackers and points of attack. (4) many motives and objectives, (5) anonymity of threats, (6) large amount of data that might be relevant to defense, (7) large noise content, (8) many possible targets, (9) availability requirements, and (10) legal constraints.

Deception may: (1) condition the target to friendly behavior, (2) divert target attention from friendly assets, (3) draw target attention to a time or place, (4) hide presence or activity from a target, (5) advertise strength or weakness as their opposites, (6) confuse or overload adversary intelligence capabilities, or (7) disguise forces.

The animal kingdom is studied briefly and characterized as ranging from concealment to simulation, at levels (1) static, (2) dynamic, (3) adaptive, and (4) premeditated.

Political science and psychological deceptions are fused into maxims; (1) pre-existing notions given excessive weight, (2) desensitization degrades vigilance, (3) generalizations or exceptions based on limited data, (4) failure to fully examine the situation limits comprehension, (5) limited time and processing power limit comprehension, (6) failure to adequately corroborate, (7) over-valuing data based on rarity, (8) experience with source may color data inappropriately, (9) focusing on a single explanation when others are available, (10) failure to consider alternative courses of action, (11) failure to adequately evaluate options, (12) failure to reconsider previously discarded possibilities, (13) ambivalence by the victim to the deception, and (14) confounding effect of inconsistent data. This is very similar to the coverage of Gilovich [Gil91] reviewed in detail elsewhere in this report.

Confidence artists use a 3-step screening process; (1) low-investment deception to gage target reaction, (2) low-risk deception to determine target pliability, and (3) reveal a deception and gage reaction to determine willingness to break the rules.

Military deception is characterized through Joint Pub 3-58 (Joint Doctrine for Military Deception) and Field Manual 90-02 [Arm98] which are already covered in this overview.

The report then goes on to review things that can be manipulated, actors, targets, contexts, and some of the then-current efforts to manipulate observables which they characterize as: (1) honeypots, (2) fishbowls, and (3) canaries. They characterize a space of (1) raw materials, (2) deception means, and (3) level of sophistication. They look at possible mission objectives of (1) shielding assets from attackers, (2) luring attention away from strategic assets, (3) the induction of noise or uncertainty, and (4) profiling identity, capabilities, and intent by creation of opportunity and observation of action. They hypothesize a deception toolkit (sic) consisting of user inputs to a rule-based system that automatically deploys deception capabilities into fielded units as needed and detail some potential rules for the operation of such a system in terms of deception means, material requirements, and sophistication. Consistency is identified as a problem, the potential for self-deception is high in such systems, and the problem of achieving adequate fidelity is reflected as it has been elsewhere.

The follow-up RAND study [GWM$^+$00] extends the previous results with a set of experiments in the effectiveness of deception against sample forces. They characterize deception as an element of "active network defense". Not surprisingly, they conclude that more elaborate deceptions are more effective, but they also find a high degree of effectiveness for select superficial deceptions against select superficial intelligence probes. They conclude, among other things, that deception can be effective in protection, counterintelligence, against cyber-reconnaissance, and to help to gather data about enemy reconnaissance. This is consistent with previous results that were more speculative. Counter deception issues are also discussed, including (1) structural, (2) strategic, (3) cognitive, (4) deceptive, and (5) overwhelming approaches.

- **Theoretical Work:** One historical and three current theoretical efforts have been undertaken in this area, and all are currently quite limited. Cohen looked at a mathematical structure of simple defensive network deceptions in 1999 [Coh99a] and concluded that as a counter-intelligence tool, network-based deceptions could be of significant value, particularly if the quality of the deceptions could be made good enough. Cohen suggested the use of rerouting methods combined with live systems of the sorts being modeled as yielding the highest fidelity in a deception. He also expressed the limits of fidelity associated with system content, traffic patterns, and user behavior, all of which could be simulated with increasing accuracy for increasing cost. In this paper, networks of up to 64,000 IP addresses were emulated for high quality deceptions using a technology called D-WALL [Coh00a].

  Dorothy Denning of Georgetown University is undertaking a small study of issues in deception. Matt Bishop of the University of California at Davis is undertaking a study funded by the Department of Energy on the mathematics of deception. Glen Sharlun of the Naval Post Graduate School is finishing a Master's thesis on the effect of deception as a deterrent and as a detection method in large-scale distributed denial of service attacks.

- **Custom Deceptions:** Custom deceptions have existed for a long time, but only recently have they gotten adequate attention to move toward high fidelity and large scales.

The reader is asked to review the previous citation [Coh99b] for more thorough coverage of computer-based defensive deceptions and to get a more complete understanding of the application of deceptions in this arena over the last 50 years.

Another major area of information protection through deception is in the area of steganography. The term steganography comes from the Greek 'steganos' (covered or secret) and 'graphy' (writing or drawing) and thus means, literally, covered writing. As commonly used today, steganography is closer to the art of information hiding, and is ancient form of deception used by everyone from ruling politicians to slaves. It has existed in one form or another for at least 2000 years, and probably a lot longer.

With the increasing use of information technology and increasing fears that information will be exposed to those it is not intended for, steganography has undergone a sort of emergence. Computer programs that automate the processes associated with digital steganography have become widespread in recent years. Steganographic content is now commonly hidden in graphic files, sound files, text files, covert channels, network packets, slack space, spread spectrum signals, and video conferencing systems. Thus steganography has become a major method for concealment in information technology and has broad applications for defense.

2.4 The Nature of Deception

Even the definition of deception is illusive. As we saw from the circular dictionary definition presented earlier, there is no end to the discussion of what is and is not deception. This not withstanding, there is an end to this paper, so we will not be making as precise a definition as we might like to. Rather, we will simply assert that:

> Deception is a set of acts that seek to increase the chances that a set of targets will behave in a desired fashion when they would be less likely to behave in that fashion if they knew of those acts.

We will generally limit our study of deceptions to targets consisting of people, animals, computers, and systems comprised of these things and their environments. While it could be argued that all deceptions of interest to warfare focus on gaining compliance of people, we have not adopted this position. Similarly, from a pragmatic viewpoint, we see no current need to try to deceive some other sort of being.

While our study will seek general understanding, our ultimate focus is on deception for information protection and is further focused on information technology and systems that depend on it. At the same time, in order for these deceptions to be effective, we have to, at least potentially, be successful at deception against computers used in attack, people who operate and program those computers, and ultimately, organizations that task those people and computers. Therefore, we must understand deception that targets people and organizations, not just computers.

2.4.1 Limited Resources lead to Controlled Focus of Attention

There appear to be some features of deception that apply to all of the targets of interest. While the detailed mechanisms underlying these features may differ, commonalities are worthy of note. Perhaps the core issue that underlies the potential for success of deception as a whole is that all targets not only have limited overall resources, but they have limited abilities to process the available sensory data they are able to receive. This leads to the notion that, in addition to controlling the set of information available to the targets, deceptions may seek to control the focus of attention of the target.

In this sense, deceptions are designed to emphasize one thing over another. In particular, they are designed to emphasize the things you want the targets to observe over the things you do not want them to observe. While many who have studied deception in the military context have emphasized the desire for total control over enemy observables, this tends to be highly resource consumptive and very difficult to do. Indeed, there is not a single case in our review of military history where such a feat has been accomplished and we doubt whether such a feat will ever be accomplished.

> *Example: Perhaps the best example of having control over observables was in the Battle of Britain in World War II when the British turned all of the Nazi intelligence operatives in Britain into double agents and combined their reports with false fires to try to get the German Air Force to miss their factories. But even this incredible level of success in deception did not prevent the Germans from creating technologies such as radio beam guidance systems that resulted in accurate targeting for periods of time.*

It is generally more desirable from an assurance standpoint to gain control over more target observables, assuming you have the resources to affect this control in a properly coordinated manner, but the reason for this may be a bit surprising. The only reason to control more observables is to increase the likelihood of attention being focused on observables you control. If you could completely control focus of attention, you would only need to control a very small number of observables to have complete effect. In addition, the cost of controlling observables tends to increase non-linearly with increased fidelity. As we try to reach perfection, the costs presumably become infinite. Therefore, there should be some cost benefit analysis undertaken in deception planning and some metrics are required in order to support such analysis.

2.4.2 All Deception is a Composition of Concealments and Simulations

Reflections of world events appear to the target as observables. In order to affect a target, we can only create causes in the world that affect those observables. Thus all deceptions stem from the ability to influence target observables. At some level, all we can do is create world events whose reflection appear to the target as observables or prevent the reflections of world events from being observed by the target. As terminology, we will call induced reflections *'simulations'* and inhibition of reflections *'concealments'*. In general then, all deceptions are formed from combinations of concealments and simulations.

Put another way, deception consists of determining what we wish the target to observe and not observe and creating simulations to induce desired observations while using concealments to inhibit undesired observations. Using the notion of focus of attention, we can create simulations and concealments by inducing focus on desired observables while drawing focus away from undesired observables. Simulation and concealment are used to affect this focus and the focus then produces more effective simulation and concealment.

2.4.3 Memory and Cognitive Structure Force Uncertainty, Predictability, and Novelty

All targets have limited memory state and are, in some ways, inflexible in their cognitive structure. While space limits memory capabilities of targets, in order to be able to make rapid and effective decisions, targets necessarily trade away some degree of flexibility. As a result, targets have some predictability. The problem at hand is figuring out how to reliably make target behavior (focus of attention, decision processes, and ultimately actions) comply with our desires. To a large extent, the purpose of this study is to find ways to increase the certainty of target compliance by creating improved deceptions.

There are some severe limits to our ability to observe target memory state and cognitive structure. Target memory state and detailed cognitive structure is almost never fully available to us. Even if it were available, we would be unable, at least at the present, to adequately process it to make detailed predictions of behavior because of the complexity of such computations and our own limits of memory and cognitive structure. This means that we are forced to make imperfect models and that we will have uncertain results for the foreseeable future.

While modeling of enough of the cognitive structures and memory state of targets to create effective deceptions may often be feasible, the more common methods used to create deceptions are the use of characteristics that have been determined through psychological studies of human behavior, animal behavior, analytical and experimental work done with computers, and psychological studies done on groups. The studies of groups containing humans and computers are very limited at and those that do exist ignore the emerging complex global network environment. Significant additional effort will be required in order to understand common modes of deception that function in the combined human-computer social environment.

A side effect of memory is the ability of targets to learn from previous deceptions. Effective deceptions must be novel or varied over time in cases where target memory affects the viability of the deception.

2.4.4 Time, Timing, and Sequence are Critical

Several issues related to time come up in deceptions. In the simplest cases, a deception might come to mind just before it is to be performed, but for any complex deception, pre-planning is required, and that pre-planning takes time. In cases where special equipment or other capabilities must be researched and developed, the entire deception process can take months to years.

In order for deception to be effective in many real-time situations, it must be very rapidly deployed. In some cases, this may mean that it can be activated almost instantaneously. In other cases this may mean a time frame of seconds to days or even weeks or months. In strategic deceptions such as those in the Cold War, this may take place over periods of years.

In every case, there is some delay between the invocation of a deception and its effect on the target. At a minimum, we may have to contend with speed of light effects, but in most cases, cognition takes from milliseconds to seconds. In cases with higher momentum, such as organizations or large systems, it may take minutes to hours before deceptions begin to take effect. Some deceptive information is even planted in the hopes that it will be discovered and acted on in months to years.

Eventually, deceptions may be discovered. In most cases a critical item to success in the deception is that the time before discovery be long enough for some other desirable thing to take place. For one-shot deceptions intended to gain momentary compliance, discovery after a few seconds may be adequate, but other deceptions require longer periods over which they must be sustained. Sustaining a deception is generally related to preventing its discovery in that, once discovered, sustainment often has very different requirements.

Finally, nontrivial deceptions involve complex sequences of acts, often involving branches based on feedback attained from the target. In almost all cases, out of the infinite set of possible situations that may arise, some set of critical criteria are developed for the deception and used to control sequencing. This is necessary because of the limits of the ability of deception planning to create sequencers for handling more complex decision processes, because of limits on available observables for feedback, and because of limited resources available for deception.

> *Example: In a commonly used magician's trick, the subject is given a secret that the magician cannot possibly know based on the circumstances. At some time in the process, the subject is told to reveal the secret to the whole audience. After the subject makes the secret known, the magician reveals that same secret from a hiding place. The trick comes from the sequence of events. As soon as the answer is revealed, the magician chooses where the revealed secret is hidden. What really happens is that the magician chooses the place based on what the secret is and reveals one of the many pre-planted secrets. If the sequence required the magician to reveal their hidden result first, this deception would not work[Fel00].*

2.4.5 Observables Limit Deception

In order for a target to be deceived, their observations must be affected. Therefore, we are limited in our ability to deceive based on what they are able to observe. Targets may also have allies with different observables and, in order to be effective, our deceptions must take those observables into account. We are limited both by what can be observed and what cannot be observed. What cannot be observed we cannot use to induce simulation, while what can be observed creates limits on our ability to do concealment.

> *Example: Dogs are commonly used in patrol units because of the fact that they have different sensory and cognitive capabilities than people have. Thus when people try to conceal themselves from other people, the things they choose to do tend to fool other people but not animals like dogs which, for example, might smell them out even without seeing or hearing them.*

Our own observables also limit our ability to do deceptions because sequencing of deceptions depends on feedback from the target and because our observables in terms of accurate intelligence information drive our ability to understand the observables of the target and the effect of those observables on the target.

2.4.6 Operational Security is a Requirement

Secrecy of some sort is fundamental to all deception, if only because the target would be less likely to behave in the desired fashion if they knew of the deception (by our definition above). This implies operational security of some sort.

One of the big questions to be addressed in some deceptions is who should be informed of the specific deceptions under way. Telling too many people increases the likelihood of the deception being leaked to the target. Telling too few people may cause the deception to fool your own side into blunders.

> *Example: In Operation Overlord during World War II, some of the allied deceptions were kept so secret that they fooled allied commanders into making mistakes. These sorts of errors can lead to fratricide* [Dew89].

Security is expensive and creates great difficulties, particularly in technology implementations. For example, if we create a device that is only effective if its existence is kept secret, we will not be able to apply it very widely, so the number of people that will be able to apply it will be very limited. If we create a device that has a set of operational modes that must be kept secret, the job is a bit easier. As we move toward a device that only needs to have it's current placement and current operating mode kept secret, we reach a situation where widespread distribution and effective use is feasible.

A vital issue in deception is the understanding of what must be kept secret and what may be revealed. If too much is revealed, the deception will not be as effective as it otherwise may have been. If too little is revealed, the deception will be less effective in the larger sense because fewer people will be able to apply it. History shows that device designs and implementations eventually leak out. That is why soundness for a cryptographic system is usually based on the assumption that only the keys are kept secret. The same principle would be well considered for use in many deception technologies.

A further consideration is the deterrent effect of widely published use of deception. The fact that high quality deceptions are in widespread use potentially deters attackers or alters their behavior because they believe that they are unable to differentiate deceptions from non-deceptions or because they believe that this differentiation substantially increases their workload. This was one of the notions behind Deception ToolKit (DTK)[Coh98a]. The suggestion was even made that if enough people use the DTK deception port, the use of the deception port alone might deter attacks.

2.4.7 Cybernetics and System Resource Limitations

In the systems theory of Norbert Weiner (called Cybernetics) [Wei48] many systems are described in terms of feedback. Feedback and control theory address the notions of systems with expectations and error signals. Our targets tend to take the difference between expected inputs and actual inputs and adjust outputs in an attempt to restore stability. This feedback mechanism both enables and limits deception.

Expectations play a key role in the susceptibility of the target to deception. If the deception presents observables that are very far outside of the normal range of expectations, it is likely to be hard for the target to ignore it. If the deception matches a known pattern, the target is likely to follow the expectations of that pattern unless there is a reason not to. If the goal is to draw attention to the deception, creating more difference is more likely to achieve this, but it will also make the target more likely to examine it more deeply and with more skepticism. If the object is to avoid something being noticed, creating less apparent deviation from expectation is more likely to achieve this.

Targets tend to have different sensitivities to different sorts and magnitudes of variations from expectations. These result from a range of factors including, but not limited to, sensor limitations, focus of attention, cognitive structure, experience, training, reasoning ability, and pre-disposition. Many of these can be measured or influenced in order to trigger or avoid different levels of assessment by the target.

Most systems do not do deep logical thinking about all situations as they arise. Rather, they match known patterns as quickly as possible and only apply the precious deep processing resources to cases where pattern matching fails to reconcile the difference between expectation and interpretation. As a result, it is often easy to deceive a system by avoiding its logical reasoning in favor of pattern matching. Increased rush, stress, uncertainty, indifference, distraction, and fatigue all lead to less thoughtful and more automatic responses in humans [Cia01]. Similarly, we can increase human reasoning by reduced rush, stress, certainty, caring, attention, and alertness.

Example: Someone who looks like a valet parking person and is standing outside of a pizza place will often get car keys from wealthy customers. If the customers really used reason, they would probably question the notion of a valet parking person at a pizza place, *but their mind is on food and conversation and perhaps they just miss it. This particular experiment was one of many done with great success by Whitlock [Whi97].*

Similar mechanisms exist in computers where, for example, we can suppress high level cognitive functions by causing driver-level response to incoming information or force high level attention and thus overwhelm reasoning by inducing conditions that lead to increased processing regimens.

2.4.8 The Recursive Nature of Deception

The interaction we have with targets in a deception is recursive in nature. To get a sense of this, consider that while we present observables to a target, the target is presenting observables to us. We can only judge the effect of our deception based on the observables we are presented with and our prior expectations influence how we interpret these observables. The target may also be trying to deceive us, in which case, they are presenting us with the observables they think we expect to see, but at the same time, we may be deceiving them by presenting the observables we expect them to expect us to present. This goes back and forth potentially without end. It is covered by the well known story:

The Russian and US ambassadors met at a dinner party and began discussing in their normal manner. When the subject came to the recent listening device, the Russian explains that they knew about it for some time. The American explains that they knew the Russians knew for quite a while. The Russian explains they they knew the Americans knew they knew. The American explains that they knew the Russians knew that the Americans knew they knew. The Russian states that they knew they knew they knew they knew they knew they knew. The American exclaims "I didn't know that!".

To handle recursion, it is generally accepted that you must first characterize what happens at a single level, including the links to recursion, but without delving into the next level those links lead to. Once your model of one level is completed, you then apply recursion without altering the single level model. We anticipate that by following this methodology we will gain efficiency and avoid mistakes in understanding deceptions. At some level, for any real system, the recursion must end for there is ground truth. The question of where it ends deals with issues of confidence in measured observables and we will largely ignore this issues throughout the remainder of this paper.

2.4.9 Large Systems are Affected by Small Changes

In many cases, a large system can be greatly affected by small changes. In the case of deception, it is normally easier to make small changes without the deception being discovered than to directly make the large changes that are desired. The indirect approach then tells us that we should try to make changes that cause the right effects and go about it in an unexpected and indirect manner.

As an example of this, in a complex system with many people, not all participants have to be affected in order to cause the system to behave differently than it might otherwise. One method for influencing an organizational decision is to categorize the members into four categories: zealots in favor, zealots opposed, neutral parties, and willing participants. The object of this influence tactic in this case is to get the right set of people into the right categories.

Example: Creating a small number of opposing zealots will stop an idea in an organization that fears controversy. Once the set of desired changes is understood, moves can be generated with the objective of causing these changes. For example, to get an opposing zealot to reduce their opposition, you might engage them in a different effort that consumes so much of their time that they can no longer fight as hard against the specific item you wish to get moved ahead.

This notion of finding the right small changes and backtracking to methods to influence them seems to be a general principle of organizational deception, but there has only been limited work on characterizing these effects at the organizational level.

2.4.10 Even Simple Deceptions are Often Quite Complex

In real attacks, things are not so simple as to involve only a single deception element against a nearly stateless system. Even relatively simple deceptions may work because of complex processes in the targets.

> *As a simple example, we analyzed a specific instance of audio surveillance, which is itself a subclass of attack mechanism called audio/video viewing. In this case, we are assuming that the attacker is exploiting a little known feature of cellular telephones that allows them to turn on and listen to conversations without alerting the targets. This is a deception because the attacker is attempting to conceal the listening activity so that the target will talk when they otherwise might not, and it is a form of concealment because it is intended to avoid detection by the target. From the standpoint of the telephone, this is a deception in the form of simulation because it involves creating inputs that cause the telephone to act in a way it would not otherwise act (presuming that it could somehow understand the difference between owner intent and attacker intent - which it likely can not). Unfortunately, this has a side effect.*
>
> *When the telephone is listening to a conversation and broadcasting it to the attacker it consumes battery power at a higher rate than when it is not broadcasting and it emits radio waves that it would otherwise not emit. The first objective of the attacker would be to have these go unnoticed by the target. This could be enhanced by selective use of the feature so as to limit the likelihood of detection, again a form of concealment.*
>
> *But suppose the target notices these side effects. In other words, the inputs do get through to the target. For example, suppose the target notices that their new batteries don't last the advertised 8 hours, but rather last only a few hours, particularly on days when there are a lot of meetings. This might lead them to various thought processes. One very good possibility is that they decide the problem is a bad battery. In this case, the target's association function is being misdirected by their predisposition to believe that batteries go bad and a lack of understanding of the potential for abuse involved in cell phones and similar technologies. The attacker might enhance this by some form of additional information if the target started becoming suspicious, and the act of listening might provide additional information to help accomplish this goal. This would then be an act of simulation directed against the decision process of the target.*
>
> *Even if the target becomes suspicious, they may not have the skills or knowledge required to be certain that they are being attacked in this way. If they come to the conclusion that they simply don't know how to figure it out, the deception is affecting their actions by not raising it to a level of priority that would force further investigation. This is a form of concealment causing them not to act.*
>
> *Finally, even if they should figure out what is taking place, there is deception in the form of concealment in that the attacker may be hard to locate because they are hiding behind the technology of cellular communication.*
>
> *But the story doesn't really end there. We can also look at the use of deception by the target as a method of defense. A wily cellular telephone user might intentionally assume they are being listened to some of the time and use deceptions to test out this proposition. The same response might be generated in cases where an initial detection has taken place. Before association to a bad battery is made, the target might decide to take*

some measurements of radio emissions. This would typically be done by a combination of concealment of the fact that the emissions were being measured and the inducement of listening by the creation of a deceptive circumstance (i.e., simulation) that is likely to cause listening to be used. The concealment in this case is used so that the target (who used to be the attacker) will not stop listening in, while the simulation is used to cause the target to act.

The complete analysis of this exchange is left as an exercise to the reader.. good luck. To quote the immortal Bard:

Oh what a tangled web we weave when first we practice to deceive

2.4.11 Simple Deceptions are Combined to Form Complex Deceptions

Large deceptions are commonly built up from smaller ones. For example, the commonly used 'big con' plan [Far98] goes something like this: find a victim, gain the victim's confidence, show the victim the money, tell the tale, deliver a sample return on investment, calculate the benefits, send the victim for more money, take them for all they have, kiss off the victim, keep the victim quiet. Of these, only the first does not require deceptions. What is particularly interesting about this very common deception sequence is that it is so complex and yet works so reliably. Those who have perfected its use have ways out at every stage to limit damage if needed and they have a wide number of variations for keeping the target (called victim here) engaged in the activity.

2.4.12 Knowledge of the Target

The intelligence requirements for deception are particularly complex to understand because, presumably, the target has the potential for using deception to fool the attacker's intelligence efforts. In addition, seemingly minor items may have a large impact on our ability to understand and predict the behavior of a target. As was pointed out earlier, intelligence is key to success in deception. But doing a successful deception requires more than just intelligence on the target. To get to high levels of surety against capable targets, it is also important to anticipate and constrain their behavioral patterns.

In the case of computer hardware and software, in theory, we can predict precise behavior by having detailed design knowledge. Complexity may be driven up by the use of large and complicated mechanisms (e.g., try to figure out why and when Microsoft Windows will next crash) and it may be very hard to get details of specific mechanisms (e.g., what specific virus will show up next). While generic deceptions (e.g., false targets for viruses) may be effective at detecting a large class of attacks, there is always an attack that will, either by design or by accident, go unnoticed (e.g., not infect the false targets). The goal of deceptions in the presence of imperfect knowledge (i.e., all real-world deceptions) is to increase the odds. The question of what techniques increase or decrease odds in any particular situation drives us toward deceptions that tend to drive up the computational complexity of differentiation between deception and non-deception for large classes of situations. This is intended to exploit the limits of available computational power by the target. The same notions can be applied to human deception. We never have perfect knowledge of a human target, but in various aspects, we can count on certain limitations. For example, overloading a human target with information will tend to make concealment more effective.

Example: One of the most effective uses of target knowledge in a large-scale deception was the deception attack against Hitler that supported the D-day invasions of World War II. Hitler was specifically targeted in such a manner that he would personally prevent the German military from responding to the Normandy invasion. He was induced not to act when he otherwise would have by a combination of deceptions that convinced him that the invasion would be at Pas de Calais. They were so effective that they continued to *work for as much as a week after troops were inland from Normandy. Hitler thought that Normandy was a feint to cover the real invasion and insisted on not moving troops to stop it.*

The knowledge involved in this grand deception came largely from the abilities to read German encrypted Enigma communications and psychologically profile Hitler. The ability to read ciphers was, of course, facilitated by other deceptions such as over attribution of defensive success to radar. Code breaking had to be kept secret to in order to prevent the changing of code mechanisms, and in order for this to be effective, radar was used as the excuse for being able to anticipate and defend against German attacks. [Kah67]

Knowledge for Concealment The specific knowledge required for effective concealment is details of detection and action thresholds for different parts of systems. For example, knowing the voltage used for changing a 0 to a 1 in a digital system leads to knowing how much additional signal can be added to a wire while still not being detected. Knowing the electromagnetic profile of target sensors leads to better understanding of the requirements for effective concealment from those sensors. Knowing how the target's doctrine dictates responses to the appearance of information on a command and control system leads to understanding how much of a profile can be presented before the next level of command will be notified. Concealment at any given level is attained by remaining below these thresholds.

Knowledge for Simulation The specific knowledge required for effective simulation is a combination of thresholds of detection, capacity for response, and predictability of response. Clearly, simulation will not work if it is not detected and therefore detection thresholds must be surpassed. Response capacity and response predictability are typically for more complex issues.

Response capacity has to do with quantity of available resources and ability to use them effectively. For computers, we know pretty well the limits of computational and storage capacity as well as what sorts of computations can be done in how much time. While clever programmers do produce astonishing results, for those with adequate understanding of the nature of computation, these results lead clearly toward the nature of the breakthrough. We constantly face deceptions, perhaps self-deceptions, in the proposals we see for artificial intelligence in computer systems and can counter it based on the understanding of resource consumption issues. Similarly, humans have limited capacity for handling situations and we can predict these limits at some level generically and in specific through experiments on individuals. Practice may allow us to build certain capacities to an artificially high level. The use of automation to augment capacities is one of the hallmarks of human society today, but even with augmentation, there are always limits.

Response predictability may be greatly facilitated by the notions of cybernetic stability. As long as we don't exceed the capacity of the system to handle change, systems designed for stability will have predictable tendencies toward returning to equilibrium. One of the great advantages of term limits on politicians, particularly at the highest levels, is that each new leader has to be recalibrated by those wishing to target them. It tends to be easier to use simulation against targets that have been in place for a long time because their stability criteria can be better measured and tested through experiment.

2.4.13 Legality

There are legal limitations on the use of deception for those who are engaged in legal activities, while those who are engaged in illegal activities, risk jail or, in some cases, death for their deceptions.

In the civilian environment, deceptions are acceptable as a general rule unless they involve a fraud, reckless endangerment, or libel of some sort. For example, you can legally lie to your wife (although I would advise against it), but if you use deception to get someone to give you money, in most cases it's called fraud and carries a possible prison sentence. You can legally create deceptions to defeat attacks against computer systems, but there are limits to what you can do without creating potential civil liability. For example, if you hide a virus in software and it is stolen and damages the person who stole it or an innocent bystander, you may be subject to civil suit. If someone is injured as a side effect, reckless endangerment may be involved.

Police and other governmental bodies have different restrictions. For example, police may be subject to administrative constraints on the use of deceptions, and in some cases, there may be a case for entrapment if deceptions are used to create crimes that otherwise would not have existed. For agencies like the CIA and NSA, deceptions may be legally limited to affect those outside the United States, while for other agencies, restrictions may require activities only within the United States. Similar legal restrictions exist in most nations for different actions by different agencies of their respective governments. International law is less clear on how governments may or may not deceive each other, but in general, governmental deception is allowed and is widely used.

Military environments also have legal restrictions, largely as a result of international treaties. In addition, there are codes of conduct for most militaries and these include requirements for certain limitations on deceptive behavior. For example, it is against the Geneva convention to use Red Cross or other similar markings in deceptions, to use the uniform of the enemy in combat (although use in select other circumstances may be acceptable), to falsely indicate a surrender as a feint, and to falsely claim there is an armistice in order to draw the enemy out. In general, there is the notion of good faith and certain situations where you are morally obligated to speak the truth. Deceptions are forbidden if they contravene any generally accepted rule or involve treachery or perfidy. It is especially forbidden to make improper use of a flag of truce, the national flag, the military insignia and uniform of the enemy, or the distinctive badges of the Geneva convention. [Arm98] Those violating these conventions risk punishment ranging up to summary execution in the field.

Legalities are somewhat complex in all cases and legal council and review should be considered before any questionable action.

2.4.14 Modeling Problems

From the field of game theory, many notions about strategic and tactical exchanges have been created. Unfortunately, game theory is not as helpful in these matters as it might be both because it requires that a model be made in order to perform analysis and because, for models as complex as the ones we are already using in deception analysis, the complexity of the resulting decision trees often become so large as to defy computational solution. Fortunately, there is at least one other way to try to meet this challenge. This solution lies in the area of "model-based situation anticipation and constraint" [Coh99c]. In this case, we use large numbers of simulations to sparsely cover a very large space.

In each of these cases, the process of analysis begins with models. Better models generally result in better results but sensitivity analysis has shown that we do not need extremely accurate models to get usable statistical results and meaningful tactical insight[Coh99c]. This sort of modeling of deception and the scientific investigation that supports accurate modeling in this area has not yet begun in earnest, but it seems certain that it must.

One of the keys to understanding deception in a context is that the deceptions are oriented toward the overall systems that are our targets. In order for us to carry out meaningful analysis, we must have meaningful models. If we do not have these models, then we will likely create a set of deceptions that succeed against the wrong targets and fail against the desired targets, and in particular, we will most likely be deceiving ourselves.

The main problem we must first address is what to model. In our case, the interest lies in building more effective deceptions to protect systems against attacks.

These targets of such defensive deceptions vary widely and they may ultimately have to be modeled in detail independently of each other, but there are some common themes. In particular, we believe we will need to build cognitive models of computer systems, humans, and their interactions as components of target systems. Limited models of attack strengths and types associated with these types of targets exist [Coh99c] in a form amenable to simulation and analysis. These have not been integrated into a deception framework and development has not been taken to the level of specific target sets based on reasonable intelligence estimates.

There have been some attempts to model deceptions before invoking them in the past. One series of examples is the series of deceptions starting with the Deception ToolKit[Coh99b], leading to the D-Wall [Coh99a], and then to the other projects. In these cases, increasingly detailed models of targets of defensive deceptions were made and increasingly complex and effective deceptions were achieved.

2.4.15 Unintended Consequences

Deceptions may have many consequences, and these may not all be intended when the deceptions are used. Planning to avoid unintended consequences and limit the effects of the deceptions to just the target raises complex issues.

> *Example: When deception was first implemented to limit the effectiveness of computer network scanning technology, one side effect was to deceive the tools used by the defenders to detect their own vulnerabilities. In order for the deceptions to work against attackers, they also had to work against the defenders who were using the same technology.*

In the case of these deception technologies, this is an intended consequence that causes defenders to become confused about their vulnerabilities. This then has to be mitigated by adjusting the results of the scanning mechanism based on knowledge of what is a known defensive deception. In general, these issues can be quite complex.

In this case, the particular problem is that the deception affected observables of cognitive systems other than the intended target. In addition the responses of the target may indirectly affect others. For example, if we force a target to spend their money on one thing, the finiteness of the resource means that they will not spend that money on something else. That something else, in a military situation, might include feeding their prisoners, who also happen to be our troops.

All deceptions have the potential for unintended consequences. From the deceiver's perspective this is then an operations security issue. If you don't tell your forces about a deception you risk it being treated as real, while telling your own forces risks revealing the deception, either through malice or the natural difference between their response to the normal situation and the known deception.

Another problem is the potential for misassociation and misattribution. For example, if you are trying to train a target to respond to a certain action on your part with a certain action or inaction on their part, the method being used for the training may be misassociated by the target so that the indicators they use are not the ones you thought they would use. In addition, as the target learns from experiencing deceptions, they may develop other behaviors that are against your desires.

2.4.16 Counterdeception

Many studies appear in the psychological literature on counterdeception [Vri00] but little work has been done on the cognitive issues surrounding computer-based deception of people and targeting computers for deception. No metrics relating to effectiveness of deception were shown in any study of computer-related deception we were able to find. The one exception is in the provisioning of computers for increased integrity, which is generally discussed in terms of (1) honesty and truthfulness, (2) freedom from unauthorized modification, and (3) correspondence to reality. Of these, only freedom from unauthorized modification has been extensively studied for computer systems. There are studies that have shown that people tend to believe what computers indicate to them, but few of these are helpful in this context.

Pamela Kalbfleisch categorized counterdeception in face-to-face interviews according to the following schema [Kal94]. (1) No nonsense, (2) Criticism, (3) Indifference, (4) Hammering, (5) Unkept secret, (6) Fait accompli, (7) Wages alone, (8) All alone, (9) Discomfort and relief, (10) Evidence bluff, (11) Imminent discovery, (12) Mum's the word (13) Encouragement, (14) Elaboration, (15) Diffusion of responsibility, (16) Just having fun, (17) Praise (18) Excuses, (19) It's not so bad, (20) Others have done worse, (21) Blaming (22) Buildup of lies, (23) No explanations allowed, (24) Repetition, (25) Compare and contrast, (26) Provocation, (27) Question inconsistencies as they appear, (28) Exaggeration, (29) Embedded discovery, (30) A chink in the defense, (31) Self-disclosure, (32) Point of deception cues, (33) You are important to me, (34) Empathy, (35) What will people think?, (36) Appeal to pride, (37) Direct approach, and (38) Silence. It is also noteworthy that most of these counterdeception techniques themselves depend on deception and stem, perhaps indirectly, from the negotiation tactics of Karrass [Kar70].

Extensive studies of the effectiveness of counter deception techniques have indicated that success rates with face-to-face techniques rarely exceed 60% accuracy and are only slightly better at identifying lies than truths. Even poorer performance result from attempts to counter deception by examining body language and facial expressions. As increasing levels of control are exerted over the subject, increasing care is taken in devising questions toward a specific goal, and increasing motivation for the subject to lie are used, the rate of deception detection can be increased with verbal techniques such as increased response time, decreased response time, too consistent or pat answers, lack of description, too ordered a presentation, and other similar indicators. The aide of a polygraph device can increase accuracy to about 80% detection of lies and more than 90% detection of truths for very well structured and specific sorts of questioning processes [Vri00].

The limits of the target in terms of detecting deception leads to limits on the need for high fidelity in deceptions. The lack of scientific studies of this issue inhibit current capabilities to make sound decisions without experimentation.

2.4.17 Summary

The dimensions and issues involved are summarized in Table 2.1.

2.5 A Model for Human Deception

By looking extensively at the literature on human cognition and deception, a model was formed of human cognition with specific focus on its application to deception. This includes Lambert's data collection and mapping into his model of human deception.

2.5.1 Lambert's Cognitive Model

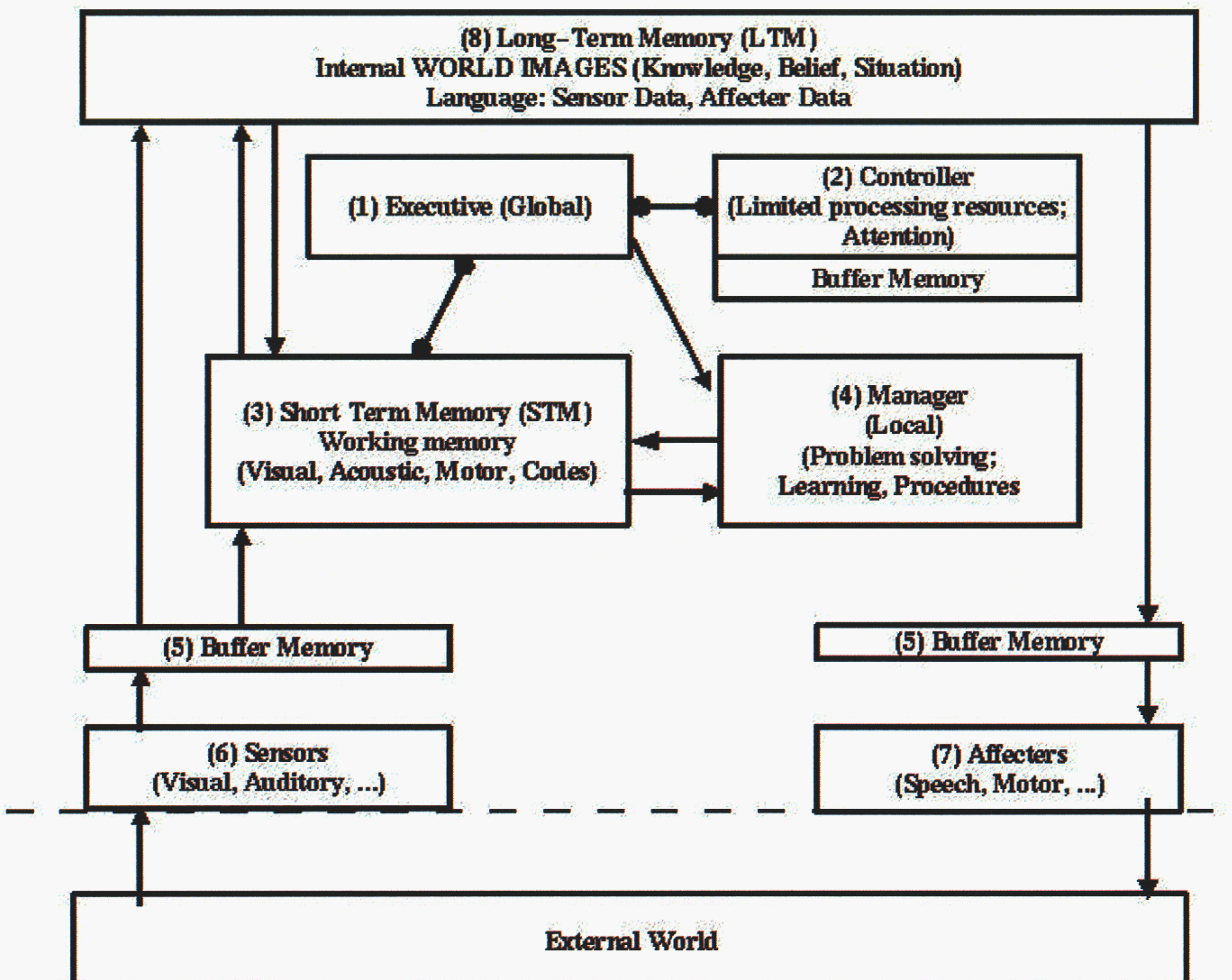
We begin with Lambert's model of human cognition [Lam87]. This model is linked to the history of psychological models of brain function and cognition and, as such, does not represent so much the physiology of the brain as the things it is generally believed to do and the manner in which it is generally believed to operate. There is no sense that this model will be found to match physiology in the long run, however, it is useful because it relates to a great deal of other experimental work that has been done on deception and the limits of human perception. It may also be related to perceptual control theory's notions of orders of control and, through that mechanistic view, to physiology [RP90].

This model shown in Figure 2.1 [2] identifies integers as labels for major brain functions. Within this model, Lambert has created a structure of sub processes identified with behavior in general and deception in particular. This structure is broken down into subsections as follows. In addition to the structural association, Lambert created a detailed mapping of how cognitive function was thought to work. The structure can be interpreted as a stimulus response network but there is

[2] An apology is in order for the quality of some of the figures in this report. They were available only as bitmap images, so suffered somewhat in the formatting process.

Limited Resources lead to Controlled Focus of Attention	By pressuring or taking advantage of pre-existing circumstances focus of attention can be stressed. In addition, focus can be inhibited, enhanced, and through the combination of these, redirected.
All Deception is a Composition of Concealments and Simulations	Concealments inhibit observation while simulations enhance observation. When used in combination they provide the means for redirection.
Memory and Cognitive Structure Force Uncertainty, Predictability, and Novelty	The limits of cognition force the use of rules of thumb as short-cuts to avoid the paralysis of analysis. This provides the means for inducing desired behavior through the discovery and exploitation of these rules of thumb in a manner that restricts or avoids higher level cognition.
Time, timing, and sequence are critical	All deceptions have limits in planning time, time to perform, time till effect, time till discovery, sustainability, and sequences of acts.
Observables Limit Deception	Target, target allies, and deceiver observables limit deception and deception control.
Operational Security is a Requirement	Determining what needs to be kept secret involves a trade off that requires metrics in order to properly address.
Cybernetics and System Resource Limitations	Natural tendencies to retain stability lead to potentially exploitable movement or retention of stability states.
The Recursive Nature of Deception	Recursion between parties leads to uncertainty that cannot be perfectly resolved but that can be approached with an appropriate basis for association to ground truth.
Large Systems are Affected by Small Changes	For organizations and other complex systems, finding the key components to move and finding ways to move them forms a tactic for the selective use of deception to great effect.
Even Simple Deceptions are Often Quite Complex	The complexity of what underlies a deception makes detailed analysis quite a substantial task.
Simple Deceptions are Combined to Form Complex Deceptions	Big deceptions are formed from small sub-deceptions and yet they can be surprisingly effective.
Knowledge of the Target	Knowledge of the target is one of the key elements in effective deception.
Legality	There are legal restrictions on some sorts of deceptions and these must be considered in any implementation.
Modeling Problems	There are many problems associated with forging and using good models of deception.
Unintended Consequences	You may fool your own forces, create mis-associations, and create mis-attributions. Collateral deception has often been observed.
Counterdeception	Target capabilities for counterdeception may result in deceptions being detected.

Table 2.1: Summary: Dimensions and Issues of Deception



System Components of the Cognitive Model

Figure 2.1: Lambert's Model of Cognition

an isomorphism to a model-referenced adaptive control system. The components consist of (1) the global executive, (2) a controller with limited processing resources and buffer memory, (3) shot-term memory and working memory which includes visual acoustic, motor, and coded memories, (4) the local manager which does problem solving, learning, and procedures, (5) buffer memories for both input and output, (6) sensors, which include transducers for the senses, (7) affecters, which includes transducers for all outputs, and (8) long-term memory, which includes internal images of the world (knowledge, belief, and situation) and language (sensor data and affector data).

The model provides for specific interconnections between components that appear to occur in humans. Specifically, long term memory is affected only by short term memory but affects short term memory and buffer memories for sensors and affectors. The executive sends information to the local manager and acts in a controlling function over short term memory and the controller. The short term memory interacts with the long-term memory, receives information from sensor buffers, and interacts with the local manager. The local manager receives information from the global executive and interacts with the short term memory. The sensor observes reflections of the world and sends the resulting signals through incoming buffer memory to short and long term memory. Long term memory feeds information to output buffers that then pass the information on to affectors.

This depiction, shown in Figure 2.2 is reflected in a different structure which models the system processes of cognition. In this depiction, we see the movement of information from senses through a cognitive process that includes reflexes, conditioned behavior, intuition, and reasoning, and a movement back down to action. Many more details are provided, but this is the general structure of cognition with which Lambert worked. From a standpoint of understanding deception, the notion is that the reflections of the world that reach the senses of the cognition system are interpreted based on its present state. The deception objective is to control those reflections so as to produce the desired changes in the perception of the target so as to achieve compliance. This can be done by inhibiting or inducing cognitive activities within this structure.

The induction of signals at the sense level is relatively obvious, and the resulting reflexive responses are quite predictable in most cases. the problems start becoming considerable as higher levels of the victim's cognitive structure get involved. while the mechanism of deception may involve the perception of feature, any feedback from this can only be seen as a result of conditioned behaviors at the perceive form level or higher level cognitive affects reflected in the ultimate drives of the system. For this reason, while the model may be helpful in understanding internal states, affects at the perceive feature level are aliased as affects at higher levels. Following the earlier depiction of deceptions as consisting of inhibitions and inducements of sensor data we can think of internal effects of deception on cognition in terms of combinations of inhibitions and inducements of internal signals. The objective of a deception might then, for example, be the inhibition of sensed content from being perceived as a feature, perhaps accomplished by a combination of reducing the available signal and distracting focus of attention by inducing the perception of a different form and causing a simultaneous reflexive action to reduce the available signal. This is precisely what is done in the case of the disappearing elephant magic trick. The disappearing elephant trick is an excellent example of the exploitation of the cognitive system and can be readily explained through Lambert's model.

Example: This trick is set up by the creation of a rippling black silk curtain behind the elephant, which is gray. The audience is in a fairly close pack staring right at the elephant some distance away. Just before the elephant disappears, a scantily clad woman walks across the front of the crowd and the magician is describing something that is not very interesting with regard to the trick. Then, as eyes turn toward the side the girl is walking toward, a loud crash sound is created to that side of the crowd. They crowd's reflexive response to a crashing sound it to turn toward the sound, which they do. This takes about 1/3 to 1/2 second. As soon as they are looking that way, the magician causes another black silk rippling curtain to rise up in front of the elephant. This takes less than 1/4 second. Because of the low contrast between the elephant and the curtain and

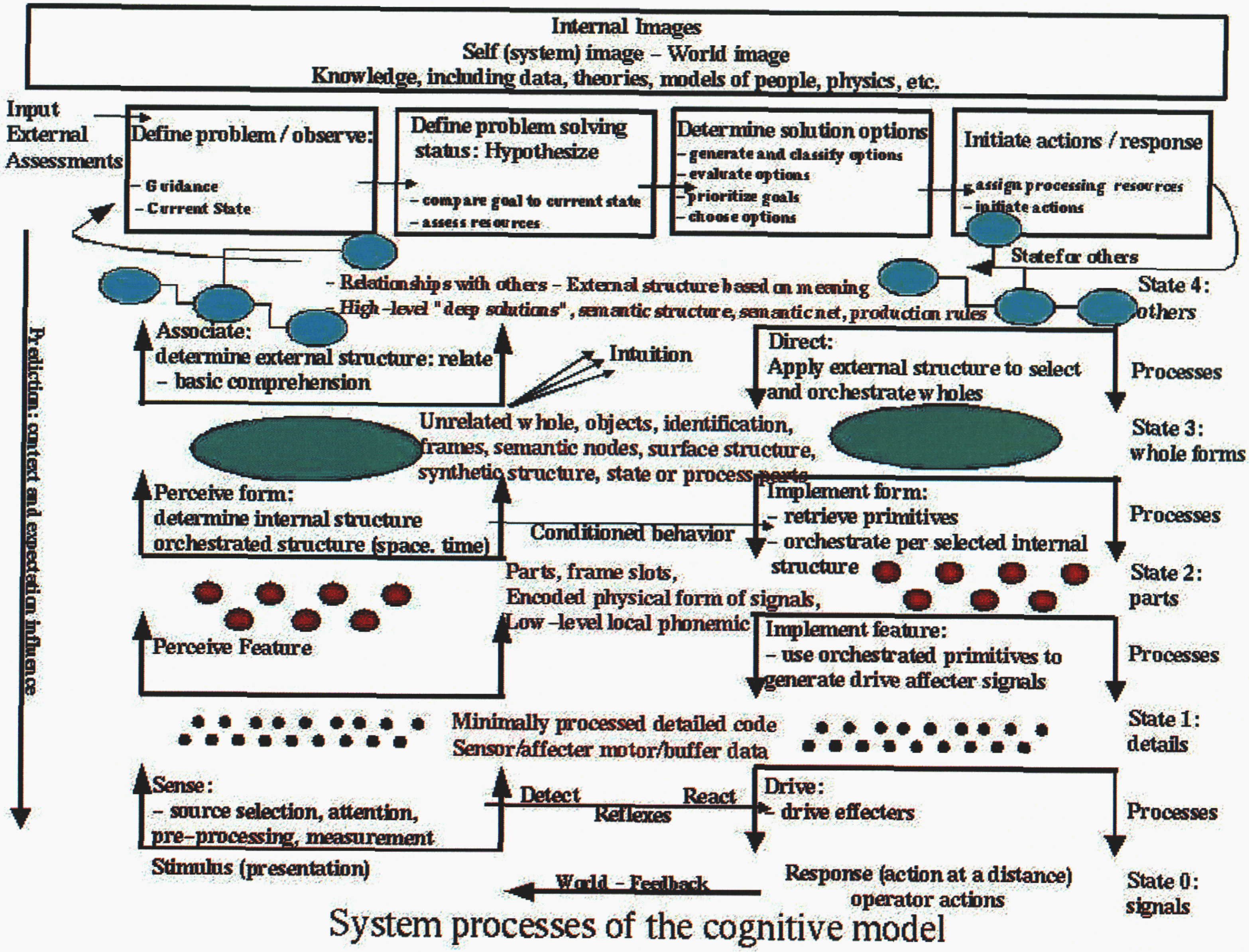


Figure 2.2: Lambert's Model of Cognition (2)

the rippling effect of the black back and front curtains, there is no edge line induced in the audience and thus attention is not pulled toward the curtains. By the time the crowd looks back, the elephant is gone and is then moved away while out of sight. The back curtain is lowered, and the front curtain is then raised to prove that only the wall remains behind the curtain.

For low-level one-step deceptions such as this one, Lambert's model is an excellent tool both for explanation and for planning. There are a set of known sensors, reflexes, and even well known or trainable conditioned responses that can be exploited almost at will. In some cases it will be necessary to force the cognitive system into a state where these prevail over higher level controlling processes, such as a member of the crowd who is focusing very carefully on what is going on. This can be done by boring them into relaxation, which the magician tries to do with his boring commentary and the more interesting scantily clad woman, but otherwise it is pretty straight forward. Unfortunately, this model provides inadequate structure for dealing with higher level or longer term cognitive deceptions. For these you need to move to another sort of model that, while still consistent with this model, provides added clarity regarding possible moves.

2.5.2 A Cognitive Model for Higher Level Deceptions

The depiction in Figure 2.3 attempts to provide additional structure for higher level cognitive deceptions. This model starts to look at how humans interact to create deceptions and how those deceptions can, at a broad level, cause interpretation and behavior in the target that is compliant with the deceiver. It also shows the recursive nature of deception because of the regress induced by both time and symmetry.

The depiction shows interaction between two human or group cognitive systems. The interaction all takes place through the world using human senses (small, taste, hearing, touching, seeing, pheromones, and allergic reactions). Deception is modeled by the induction or suppression of target observables by the deceiver.

> Cognitive processes responding directly to inputs include sensory data which, after sensor bias and the filter of a set of observables, becomes observable. Sensory data, after bias, can trigger reflexive responses which also induce observable internal changes. Other actions can also be generated and expectations actively control everything in this list. Focus of attention can also be affected at this level because of detection mechanisms and their triggering of higher level processes. This paragraph summarizes what we will tentatively call the 'low level' cognitive system.
>
> Cognitive processes in, what we tentatively call, the middle level of cognition include conditioned and other automatic but non reflexive responses, measurement mechanisms and automatic or trained evaluation and decision methods, learned and nearly automated capabilities including skills, tools, and methods that are based on pattern matching, training, instinctual responses, the actions they trigger, and the feedback mechanisms involved in controlling those actions. This level also involves learned patterns of focus of attention.
>
> The remaining cognitive processes are called high level. This includes reason-based assessments and capabilities, expectations, which include biases, fidelity of interest, level of effort, consistency with observables, and high-level focus of attention, and intent, which includes objectives, qualitative evaluation, schedule and budgetary requirements. The link between expectations and the rest of the cognitive structure is particularly important because expectations alter focus of attention sequences, cognitive biases, assessment, intent, and the evaluation of expectations, while changing of expectation can keep them stable, move them at a limited rate, or cause dissonance.

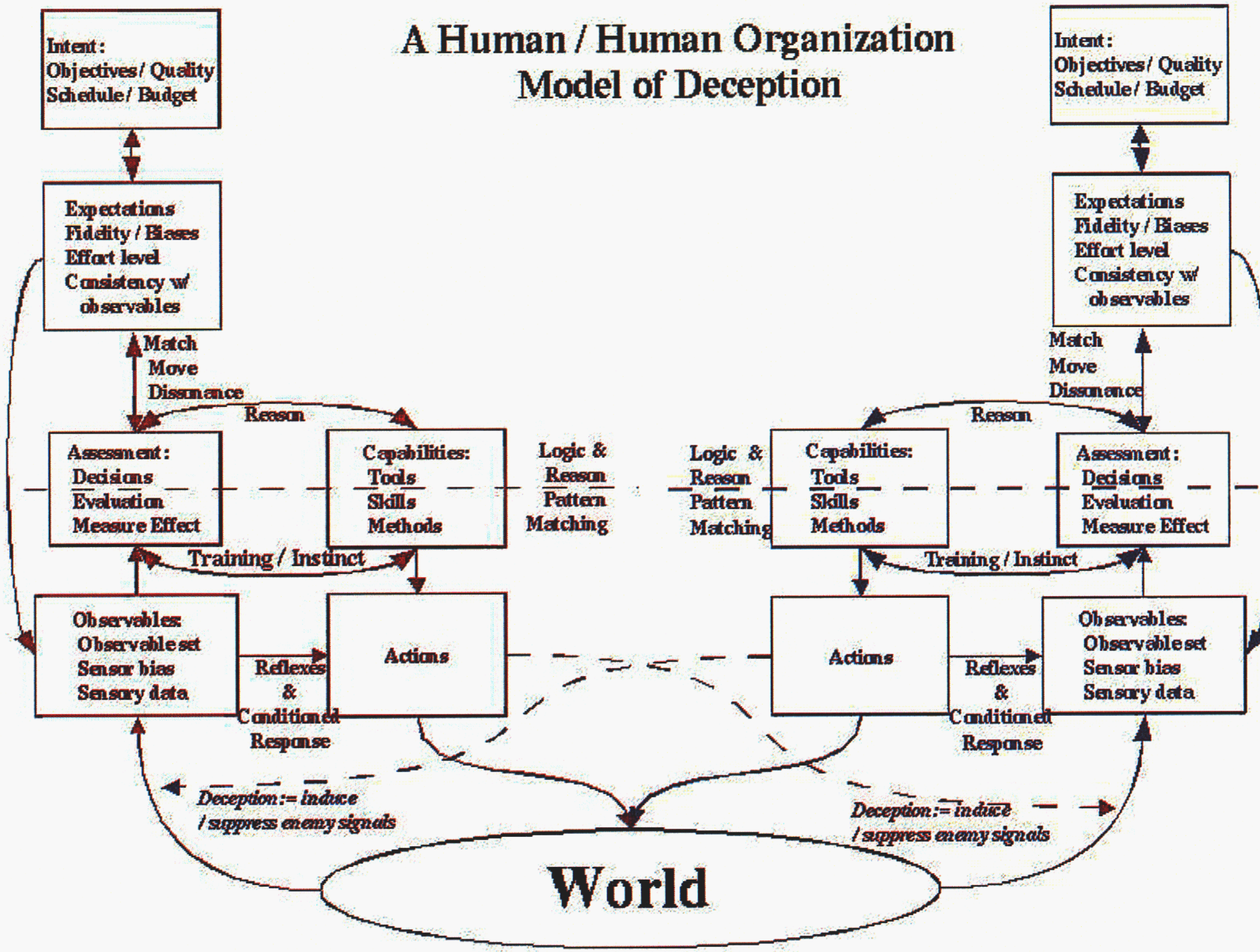


Figure 2.3: Model of Human Cognitaion for Deceptions

2.5.3 Deceptions of Low-level Cognition

In this model, we have collapsed the lower levels (up to conditioned response) of Lambert's model into the bottom two boxes (Observables and Actions) and created a somewhat more specific higher level structure. Details of these deceptions are provided in the sections 6 and 7 of Lambert's data collection. Low-level visual deceptions are demonstrated by Seckel [Sec00] and described by Hoffman [Hof98]. Audio deceptions are demonstrated on an audio CD-ROM by Deutsch [Deu95].

2.5.4 Deceptions of Mid-level Cognition

The notion is that there are pattern matching and reason-based assessments and capabilities that interact to induce more thoughtful decisions than conditioned response. While pattern matching cognition mechanisms are more thoughtful than conditioned response, they are essentially the programmed behaviors identified by Cialdini [Cia01] and some of the negotiation tactics of Karrass [Kar70]. These include, but are not limited to, reciprocation, authority, contrast, commitment and consistency, automaticity, social proof, liking, and scarcity, and as Karrass formulates it, credibility, message content and appeal, situation setting and rewards, and media choice are all methods.

The potential for decisions to be moved to more logical reasoning exists, but this is limited by the effects identified by Gilovich [Gil91]. Specifically, the notions that people (erroneously) believe that effects should resemble their causes, they misperceive random events, they misinterpret incomplete or unrepresentative data, they form biased evaluations of ambiguous and inconsistent data, they have motivational determinants of belief, they bias second hand information, and they have exaggerated impressions of social support. More content is provided in the sections numbered 1, 2, and some portions of 4 and 8 of Lambert's data collection.

2.5.5 Deceptions of High-level Cognition

Karrass [Kar70] also provides techniques for affecting influence in high-level thoughtful situations. He explains that change comes from learning and acceptance. Learning comes from hearing and understanding, while acceptance comes from comfort with the message, relevance, and good feelings toward the underlying idea. These are both affected by audience motives and values, the information and language used for presentation, audience attitudes and emotions, and the audience's perception and role in the negotiation. Karrrass provides a three dimensional depiction of goals, needs, and perceptions and asserts that people are predictable. He also provides a set of tactics including timing, inspection, authority, association, amount, brotherhood, and detour that can be applied in a deception context. Handy also provides a set of influence tactics that tend to be most useful at higher levels of reasoning, including physicality, resources, position (which yields information, access, and right to organize), expertise, personal charisma, and emotion. More content is also provided in the sections 4 and 8 of Lambert's data collection.

2.5.6 Moving from High-Level to Mid-level Cognition

Karrass also augments Cialdini's notions [Cia01] of rush, stress, uncertainty, indifference, distraction, and fatigue leading to less thoughtful and more automatic responses and brings out Maslow's needs hierarchy (basic survival, safety, love, self worth, and self-actualization). By forcing earlier sets of these issues, reasoning can be driven away and replaced by increased automaticity. Tactics of timing can also be used to drive people toward increased automaticity. Thus we can either drive the target toward less thought or use Karrass's methods of negotiation to cause desired change.

2.5.7 Moving from Mid-Level to High-level Cognition

Cognition moves to higher levels only when there are intent-based forcing factors that lead to deeper analysis, (e.g., when objectives are oriented toward more in-depth thought, quality requirements drive more detailed consideration, schedule availability provides free time to do deeper consideration, or extra budget is available for this purpose) or when expectations are not met (i.e., the fidelity of the deception is inadequate, biases trigger more detailed examination, inconsistencies or errors are above some threshold, or the difference between expectations and observations is so great or changing at so great a rate as to cause dissonance). In these cases, higher levels of reasoning are applied, complete with all of their potential logical fallacies and their special skills, tools, and methods. Higher level reasoning is desired when we wish to change intent or make radical changes in expectations, while we try to drive decisions to lower cognitive levels when we can induce less thoughtful responses in our favor.

2.5.8 An Example

To get a sense of how the model might be applied to deceptions, we have included a sample analysis of a simple human deception. The deception is an attack with a guard at a gate as the target. It happens many times each day and is commonly called tailgating.

> *The target of this deception is the guard and our method will be to try to exploit a natural overload that takes place during the return from lunch hour on a Thursday. We choose the end of the lunch hour on Thursday because the guard will be as busy as they ever get and because they will be looking forward to the weekend and will probably have a somewhat reduced alertness level. Thus we are intentionally trying to keep processing at a pattern matching level by increased rush, stress, indifference, distraction, and fatigue.*
>
> *We stand casually out of the guard's sight before the crowd comes along, join the crowd as it approaches the entry, hold a notepad where a badge appears on other peoples' attire, and stay away from the guard's side of the group. Our clothing and appearance is such that it avoids dissonance with the guard's expectations and does not affect the guard's intent in any obvious way.*
>
> *We tag along in the third row back near someone that looks generally like us and, when the guard is checking one of the other people, we ease our way over to the other side of the guard, appearing to be in the already checked group. Here we are using automaticity and social proof against the guard and liking by similarity against the group we are tailgaiting with. We are also using similarity to avoid triggering sensory detection and indifference, distraction and fatigue to avoid triggering higher level cognition.*
>
> *As the group proceeds, so do we. After getting beyond the guard's sight, we move to the back of the group and drop out as they round a corner. Here we are using automaticity, liking, and social proof against the group to go along with them, followed by moving slowly out of their notice which exploits slow movement of expectations followed by concealment from observation.*

Team members have used variations on this entry technique in red teaming exercises against facilities from time to time and have been almost universally successful in its use. It is widely published and well known to be effective. It is clearly a deception because if the guard knew you were trying to get past without a badge or authorization they would not permit the entry. While the people who use it don't typically go through this analytical process at a conscious level, they do some part of it at some level and we postulate that this is why they succeed at it so frequently.

As an aside, there should always be a backup plan for such deceptions. The typical tailgaiter, if detected, will act lost and ask the guard how to get to some building or office, perhaps finding out that this is the wrong address in the process. This again exploits elements of the deception framework designed to move the guard away from high level cognition and toward automaticity that would favor letting the attacker go and not reporting the incident.

In the control system isomorphism, we can consider this same structure as attempting to maintain internal consistency and allow change only at a limited rate. The high level control system is essentially oblivious to anything unless change happens at too high a rate or deviations of high level signals from expectations are too high. Similarly, the middle levels operate using Cialdini's rules of thumb unless a disturbance at a lower level prompts obvious dissonance and low-level control decisions (e.g., remain balanced) don't get above the reflexive and conditioned response levels unless their is a control system failure.

2.6 A Model for Computer Deception

In looking at computer deceptions it is fundamental to understand that the computer is an automaton. Anthropomorphising it into an intelligent being is a mistake in this context - a self-deception. Fundamentally, deceptions must cause systems to do things differently based on their lack of ability to differentiate deception from a non-deception. Computers cannot really yet be called 'aware' in the sense of people. Therefore, when we use a deception against a computer we are really using a deception against the skills of the human(s) that design, program, and use the computer.

In many ways computers could be better at detecting deceptions than people because of their tremendous logical analysis capability and the fact that the logical processes used by computers are normally quite different than the processes used by people. This provides some level of redundancy and, in general, redundancy is a way to defeat corruption. Fortunately for those of us looking to do defensive deception against automated systems, most of the designers of modern attack technology have a tendency to minimize their programming effort and thus tend not to include a lot of redundancy in their analysis.

People use shortcuts in their programs just as they use shortcuts their thinking. Their goal is to get to an answer quickly and in many cases without adequate information to make definitive selections. Computer power and memory are limited just like human brain power and memory are limited. In order to make efficient use of resources, people write programs that jump to premature conclusions and fail to completely verify content. In addition, people who observe computer output have a tendency to believe it. Therefore, if we can deceive the automation used by people to make decisions, we may often be able to deceive the users and avoid in-depth analysis.

Our model for computer deception starts with Cohen's "Structure of Intrusion and Intrusion Detection" [Coh00b]. In this model, a computer system and its vulnerabilities are described in terms of intrusions at the hardware, device driver, protocol, operating system, library and support function, application, recursive language, and meaning vs. content levels. The levels are all able to interact, but they usually interact hierarchically with each level interacting with the ones just above and below it. This model is depicted in Figure 2.4.

This model is based on the notion that at every level of the computer's cognitive hierarchy signals can either be induced or inhibited. The normal process is shown in black, while inhibitions are shown as grey'ed out signals, and induced signals are shown in red. All of these effect memory states and processor activities at other, typically adjacent, levels of the cognitive system. Deception detection and response capabilities are key issues in the ability to defend against deceptions so there is a concentration on the limits of detection in the following discussions.

2.6.1 Hardware Level Deceptions

If the hardware of a system or network is altered, it may behave arbitrarily differently than expected. While there is a great deal of history of tamper-detection mechanisms for physical systems, no such mechanism is or likely ever will be perfect. The use of intrusion detection systems for detecting improper modifications to hardware today consist primarily of built-in self-test mechanisms such as the power on self test (POST) routine in a typical personal computer (PC). These mechanisms are designed to detect specific sorts of random stochastic fault types and are not designed to detect malicious alterations. Thus deception of these mechanisms is fairly easy to do without otherwise altering their value in detecting fault types they already detect.

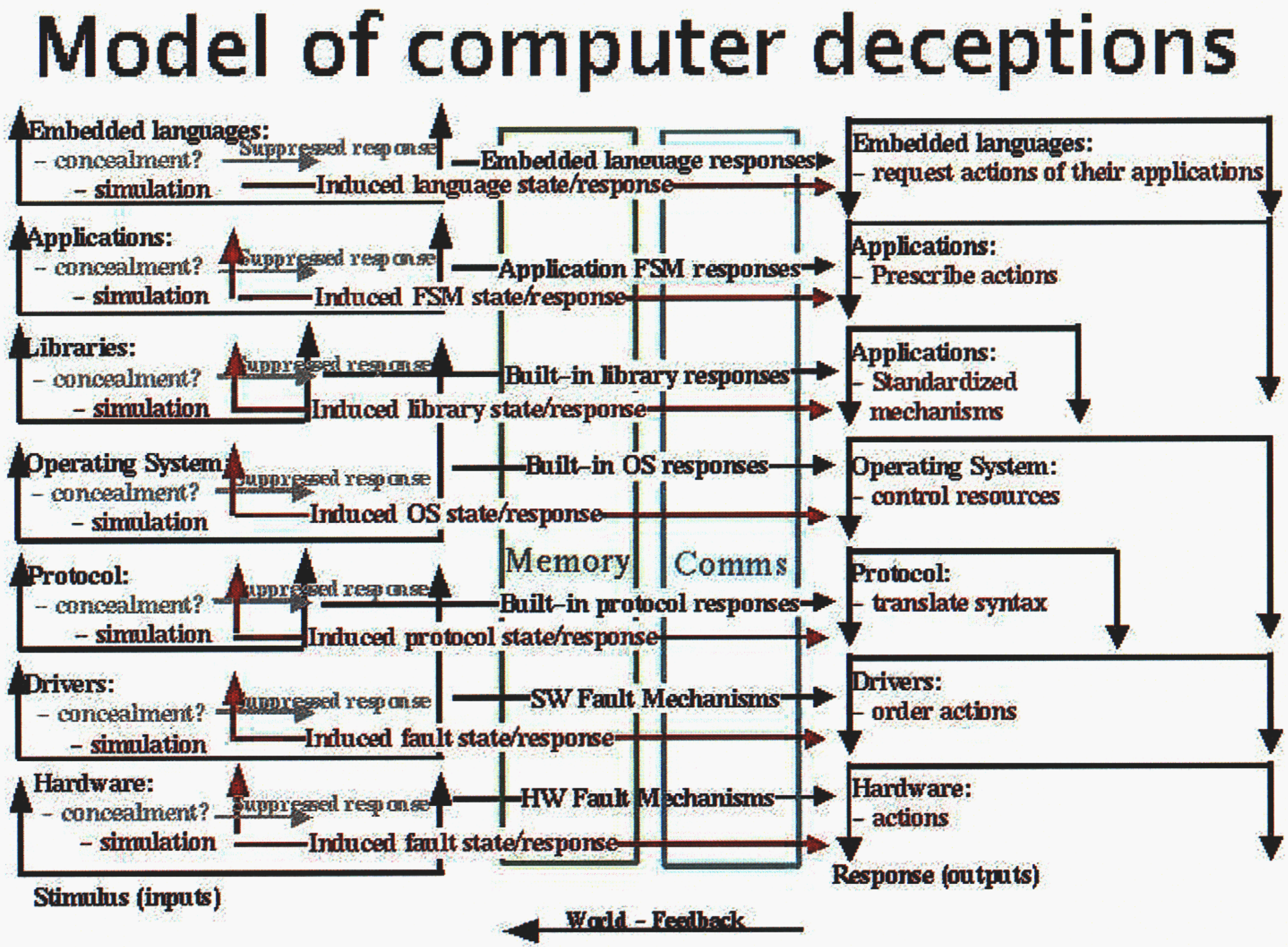


Figure 2.4: Model of Computer Cognition with deceptions

Clearly, if the hardware is altered by a serious intruder, this sort of test will not be revealing. Motion sensors, physical seals of different sorts, and even devices that examine the physical characteristics of other devices are all examples of intrusion detection techniques that may work at this level. In software, we may detect alterations in external behavior due to hardware modification, but this is only effective in large scale alterations such as the implanting of additional infrastructure. This is also likely to be ignored in most modern systems because intervening infrastructure is rarely known or characterized as part of intrusion detection and operating environments are intentionally designed to abstract details of the hardware.

Intrusions can also be the result of the interaction of hardware of different sorts rather than the specific use of a particular type of hardware. This type of intrusion mechanism appears to be well beyond the capability of current technology to detect or analyze. Deceptions exploiting these interactions will therefore likely go undetected for extended period of time. Hardware-level deceptions designed to induce desired observables are relatively easy to create and hard to detect. Induction of signals requires only knowledge of protocol and proper design of devices.

The problem with using hardware level deception for defense against serious threat types is that it requires physical access to the target system or logical access with capabilities to alter hardware level functions (e.g., microcode access). This tends to be difficult to attain against intelligence targets, if attempted against insiders it introduces deceptions that could be used against the defenders, and in the case of overrun, it does not seem feasible. That is not to say that we cannot use deceptions that operate at the hardware level against systems, but rather that affecting their hardware level is likely to be infeasible.

2.6.2 Driver Level Deceptions

Drivers are typically ignored by intrusion detection and other security systems. They are rarely inspected, in modern operating systems they can often be installed from or by applications, and they usually have unlimited hardware access. This makes them prime candidates for exploitations of all sorts, including deceptions.

A typical driver level deception would cause the driver to process items of interest without passing information to other parts of the operating environment or to exfiltrate information without allowing the system to notice that this activity was happening. It would be easy for the driver to cause widespread corruption of arbitrary other elements of the system as well as inhibiting the system from seeing undesired content.

From a standpoint of defensive deceptions, drivers are very good target candidates. A typical scenario is to require that a particular driver be installed in order to gain access to defended sites. This is commonly done with applications like RealAudio. Once the target loads the required driver, hardware level access is granted and arbitrary exploits can be launched. This technique is offensive in nature and may violate rules of engagement in a military setting or induce civil or criminal liability in a civilian setting. Its use for defensive purposes may be overly aggressive.

2.6.3 Protocol Level Deceptions

Many protocol intrusions have been demonstrated, ranging from exploitations of flaws in the IP protocol suite to flaws in cryptographic protocols. Except for a small list of known flaws that are part of active exploitations, most current intrusion detection systems do not detect such vulnerabilities. In order to fully cover such attacks, it would likely be necessary for such a system to examine and model the entire network state and effects of all packets and be able to differentiate between acceptable and unacceptable packets.

Although this might be feasible in some circumstances, the more common approach is to differentiate between protocols that are allowed and those that are not. Increasing granularity can be used to differentiate based on location, time, protocol type, packet size and makeup, and other protocol-level information. This can be done today at the level of single packets, or in some circumstances, limited sequences of packets, but it is not feasible for the combinations of packets that come from different sources and might interact within the end systems. Large scale effects can sometimes be detected, such as aggregate bandwidth utilization, but without a good model of what is supposed to happen, there will always be malicious protocol sequences that go undetected. There are also interactions between hardware and protocols. For example, there may be an exploitation of a particular hardware device which is susceptible to a particular protocol state transition, resulting in a subtle alteration to normal timing behaviors. This might then be used to exfiltrate information based on any number of factors, including very subtle covert channels.

Defensive protocol level deceptions have proven relatively easy to develop and hard to defeat. Deception ToolKit [Coh99b] and D-WALL [Coh99a] both use protocol level deceptions to great effect and these are relatively simplistic mechanisms compared to what could be devised with substantial time and effort. This appears to be a ripe area for further work. Most intelligence gathering today starts at the protocol level, overrun situations almost universally result in communication with other systems at the protocol level, and insiders generally access other systems in the environment through the protocol level.

2.6.4 Operating System Level Deceptions

At the operating system (OS) level, there are a very large number of intrusions possible, and not all of them come from packets that come over networks. Users can circumvent operating system protection in a wide variety of ways. For a successful intrusion detection system to work, it has to detect this before the attacker gains the access necessary to disable the intrusion detection mechanisms (the sensors, fusion, analysis, or response elements or the links between them can be defeated to avoid successful detection). In the late 1980s a lot of work was done in the limitations of the ability of systems to protect themselves and integrity-based self defense mechanisms were implemented that could do a reasonable job of detecting alterations to operating systems [LLN96]. These systems are not capable of defeating attacks that invade the operating system without altering files and reenter the operating system from another level after the system is functioning. Process-based intrusion detection has also been implemented with limited success. Thus we see that operating system level deceptions are commonplace and difficult to defend against.

Any host-based IDS and the analytical part of any network-based IDS involves some sort of operating environment that may be defeatable. But even if defeat is not directly attainable, denial of services against the components of the IDS can defeat many IDS mechanisms, replay attacks may defeat keep-alive protocols used to counter these denial of service attacks, selective denial of service against only desired detections are often possible, and the list goes on and on. If the operating systems are not secure, the IDS has to win a battle of time in order to be effective at detecting things it is designed to detect. Thus we see that the induction or suppression of signals into the IDS can be used to enhance or cover operating system level deceptions that might otherwise be detected.

Operating systems can have complex interactions with other operating systems in the environment as well as between the different programs operating within the OS environment. For example, variations in the timing of two processes might cause race conditions that are extremely rare but which can be induced through timing of otherwise valid external factors. Heavy usage periods may increase the likelihood of such subtle interactions, and thus the same methods that would not work under test conditions may be inducible in live systems during periods of high load. An IDS would have to detect this condition and, of course, because of the high load the IDS would be contributing to the load as well as susceptible to the effects of the attack. A specific example is the loading of a system to the point where there are no available file handles in the system tables. At this point, the IDS may not be able to open the necessary communications channels to detect, record, analyze, or respond to an intrusion.

Operating systems may also have complex interactions with protocols and hardware conditions, and these interactions are extremely complex to analyze. To date, nobody has produced an analysis of such interactions as far as we are aware. Thus deceptions based on mixed levels including the OS are likely to be undetected as deceptions.

Of course an IDS cannot detect all of the possible OS attacks. There are systems which can detect known attacks, detect anomalous behavior by select programs, and so forth, but again, a follow-up investigation is required in order for these methods to be effective, and a potentially infinite number of attacks exist that do not trip anomaly detection methods. If the environment can be characterized closely enough, it may be feasible to detect the vast majority of these attacks, but even if you could do this perfectly, there is then the library and support function level intrusion that must be addressed.

Operating systems are the most common point of attack against systems today largely because they afford a tremendous amount of cover and capability. They provide cover because of their enormous complexity and capability. They have unlimited access within the system and the ability to control the hardware so as to yield arbitrary external effects and observables. They try to control access to themselves, and thus higher level programs do not have the opportunity to measure them for the presence of deceptions. They also seek to protect themselves from the outside world so that external assessment is blocked. While they are not perfect at either of these types of protection, they are effective against the rest of the cognitive system they support. As a location for deception, they are thus prime candidates.

To use defensive deception at the target's operating system level requires offensive actions on the part of the deceiver and yields only indirect control over the target's cognitive capability. This has to then be exploited in order to affect deceptions at other levels and this exploitation may be very complex depending on the specific objective of the deception.

2.6.5 Library and Support Function Level Intrusions

Libraries and support functions are often embedded within a system and are largely hidden from the programmer so that their role is not as apparent as either operating system calls or application level programs. A good example of this is in languages like C wherein the language has embedded sets of functions that are provided to automate many of the functions that would otherwise have to be written by programmers. For example the C strings library includes a wide range of widely used functions. Unfortunately, the implementations of these functions are not standardized and often contain errors that become embedded in every program in the environment that uses them. Library-level intrusion detection has not been demonstrated at this time other than by the change detection methodology supported by the integrity-based systems of the late 1980s and behavioral detection at the operating system level. Most of the IDS mechanisms themselves depend on libraries.

An excellent recent example is the use of leading zeros in numerical values in some Unix systems. On one system call, the string -08 produces an error, while in another it is translated into the integer -8. This was traced to a library function that is very widely used. It was tested on a wide range of systems with different results on different versions of libraries in different operating environments. These libraries are so deeply embedded in operating environments and so transparent to most programmers that minor changes may have disastrous effects on system integrity and produce enormous opportunities for exploitation. Libraries are almost universally delivered in loadable form only so that source codes are only available through considerable effort. Trojan horses, simple errors, or system-to-system differences in libraries can make even the most well written and secure applications an opportunity for exploitation. This includes system applications, commonly considered part of the operating system, service applications such as web servers, accounting systems, and databases, and user level applications including custom programs and host-based intrusion detection systems.

The high level of interaction of libraries is a symptom of the general intrusion detection problem. Libraries sometimes interact directly with hardware, such as the libraries that are commonly used in special device functions like writing CD-rewritable disks. In many modern operating systems, libraries can be loaded as parts of device drivers that become embedded in the operating system itself at the hardware control level. A hardware device with a subtle interaction with a library function can be exploited in an intrusion, and the notion that any modern IDS would be able to detect this is highly suspect. While some IDS systems might detect some of the effects of this sort of attack, the underlying loss of trust in the operating environments resulting from such an embedded corruption is plainly outside of the structure of intrusion detection used today.

Using library functions for defensive deceptions offers great opportunity but, like operating systems, there are limits to the effectiveness of libraries because they are at a level below that used by higher level cognitive functions and thus there is great complexity in producing just the right effects without providing obvious evidence that something is not right.

2.6.6 Application Level Deceptions

Applications provide many new opportunities for deceptions. The apparent user interface languages offer syntax and semantics that may be exploited while the actual user interface languages may differ from the apparent languages because of programming errors, back doors, and unanticipated interactions. Internal semantics may be in error, may fail to take all possible situations into account, or there may be interactions with other programs in the environment or with state information held by the operating environment. They always trust the data they receive so that false content is easily generated and efficient. These include most intelligence tools, exploits, and other tools and techniques used by severe threats. Known attack detection tools and anomaly detection have been applied at the application level with limited success. Network detection mechanisms also tend to operate at the application level for select known application vulnerabilities.

As in every other level, there may be interactions across levels. The interaction of an application program with a library may allow a remote user to generate a complex set of interactions causing unexpected values to appear in inter-program calls, within programs, or within the operating system itself. It is most common for programmers to assume that system calls and library calls will not produce errors, and most programming environments are poor at handling all possible errors. If the programmer misses a single exception - even one that is not documented because it results from an undiscovered error in an interaction that was not anticipated - the application program may halt unexpectedly, produce incorrect results, pass incorrect information to another application, or enter an inconsistent internal state. This may be under the control of a remote attacker who has analyzed or planned such an interaction. Modern intrusion detection systems are not prepared to detect this sort of interaction.

Application level defensive deceptions are very likely to be a major area of interest because applications tend to be driven more by time to market than by surety and because applications tend to directly influence the decision processes made by attackers. For example, a defensive deception would typically cause a network scanner to make wrong decisions and report wrong results to the intelligence operative using it. Similarly, an application level deception might be used to cause a system that is overrun to act on the wrong data. For systems administrators the problem is somewhat more complex and it is less likely that application-level deceptions will work against them.

2.6.7 Recursive Languages in the Operating Environment

In many cases, application programs encode Turing Machine capable embedded languages, such as a language interpreter. Examples include Java, Basic, Lisp, APL, and Word Macros. If these languages can interpret user-level programs, there is an unlimited possible set of embedded languages that can be devised by the user or anybody the user trusts. Clearly an intrusion detection system cannot anticipate all possible errors and interactions in this recursive set of languages. This is an undecidable problem that no IDS will ever likely be able to address. Current IDS systems only address this to the extent that anomaly detection may detect changes in the behavior of the underlying application, but this is unlikely to be effective.

These recursive languages have the potential to create subtle interactions with all other levels of the environment. For example, such a language could consume excessive resources, use a graphical interface to make it appear as if it were no longer operating while actually interpreting all user input and mediating all user output, test out a wide range of known language and library interactions until it found an exploitable error, and on and on. The possibilities are literally endless. All attempts to use language constructs to defeat such attacks have failed to date, and even if they were to succeed to a limited extent, any success in this area would not be due to intrusion detection capabilities.

It seems that no intrusion detection system will ever have a serious hope of detecting errors induced at these recursive language levels as long as we continue to have user-defined languages that we trust to make decisions affecting substantial value. Unless the IDS is able to 'understand' the semantics of every level of the implementation and make determinations that differentiate desirable intent from malicious intent, the IDS cannot hope to mediate decisions that have implications on resulting values. This is clearly impossible,

Recursive languages are used in many applications including many intelligence and systems administration applications. In cases where this can be defined or understood or cases where the recursive language itself acts as the application, deceptions against these recursive languages should work in much the same manner as deceptions against the applications themselves.

2.6.8 The Meaning of the Content versus Realities

Content is generally associated with meaning in any meaningful application. The correspondence between content and realities of the world cannot reasonably be tracked by an intrusion detection system, is rarely tracked by applications, and cannot practically be tracked by other levels of the system structure because it is highly dependent on the semantics of the application that interprets it. Deceptions often involve generating human misperceptions or causing people to do the wrong thing based on what they see at the user interface. In the end, if this wrong thing corresponds to a making a different decision than is supposed to be made, but still a decision that is a feasible and reasonable one in a slightly different context, only somebody capable of making the judgment independently has any hope of detecting the error.

Only certain sorts of input redundancy are known to be capable of detecting this sort of intrusion and this becomes cost prohibitive in any large-scale operation. This sort of detection is used in some high surety critical applications, but not in most intelligence applications, most overrun situations, or by most systems administrators. The programmers of these systems call this "defensive programming" or some such thing and tend to fight against its use.

Attackers commonly use what they call 'social engineering' (a.k.a., perception management) to cause the human operator to do the wrong thing. Of course such behavioral changes can ripple through the system as well, ranging from entering wrong data to changing application level parameters to providing system passwords to loading new software updates from a web site to changing a hardware setting. All of the other levels are potentially affected by this sort of interaction.

Ultimately, deception in information systems intended to affect other systems or people will cause results at this level and thus all deceptions of this sort are well served to consider this level in their assessments.

2.6.9 Commentary

Unlike people, computers don't typically have ego, but they do have built-in expectations and in some cases automatically seek to attain 'goals'. If those expectations and goals can be met or encouraged while carrying out the deception, the computers will fall prey just as people do.

In order to be very successful at defeating computers through deception, there are three basic approaches. One approach is to create as high a fidelity deception as you can and hope that the computer will be fooled. Another is to understand what data the computer is collecting and how it analyzes the data provided to it. The third it to alter the function of the computer to comply with your needs. The high fidelity approach can be quite expensive but should not be abandoned out of hand. At the same time, the approach of understanding enemy tools can never be done definitively without a tremendous intelligence capability. The modification of cognition approach requires an offensive capability that is not always available and is quite often illegal, but all three avenues appear to be worth pursuing.

- **High Fidelity:** High fidelity deception of computers with regard to their assessment, analysis, and use against other computers tends to be fairly easy to accomplish today using tools like D-WALL [Coh99a] and the IR effort associated with this project. D-WALL created high fidelity deception by rerouting attacks toward substitute systems. The IR does a very similar process in some of its modes of operation. The notion is that by providing a real system to attack, the attacker is suitably entertained. While this is effective in the generic sense, for specific systems, additional effort must be made to create the internal system conditions indicative of the desired deception environment. This can be quite costly. These deceptions tend to operate at a protocol level and are augmented by other technologies to effect other levels of deception.

- **Defeating Specific Tools:** Many specific tools are defeated by specific deception techniques. For example, nmap and similar scans of a network seeking out services to exploit are easily defeated by tools like the Deception ToolKit [Coh99b]. More specific attack tools such as Back Orafice (BO) can be directly countered by specific emulators such as "NoBO" - a PC-based tool that emulates a system that has already been subverted with BO. Some deception systems work against substantial classes of attack tools.

- **Modifying Function:** Modifying the function of computers is relatively easy to do and is commonly used in attacks. The question of legality aside, the technical aspects of modifying function for defense falls into the area of counterattack and is thus not a purely defensive operation. The basic plan is to gain access, expand privileges, induce desired changes for ultimate compliance, leave those changes in place, periodically verify proper operation, and exploit as desired. In some cases privileges gained in one system are used to attack other systems as well. Modified function is particularly useful for getting feedback on target cognition.

The intelligence requirements of defeating specific tools may be severe, but the extremely low cost of such defenses makes them appealing. Against off-the-Internet attack tools, these defenses are commonly effective and, at a minimum, increase the cost of attack far more than they affect the cost of defense. Unfortunately, for more severe threats, such as insiders, overrun situations, and intelligence organizations, these defenses are often inadequate. They are almost certain to be detected and avoided by an attacker with skills and access of this sort. Nevertheless, from a standpoint of defeating the automation used by these types of attackers, relatively low-level deceptions have proven effective. In the case of modifying target systems, the problems become more severe in the case of more severe threats. Insiders are using your systems, so modifying them to allow for deception allows for self-deception and enemy deception of you. For overrun conditions you rarely have access to the target system, so unless you can do very rapid and automated modification, this tactic will likely fail. For intelligence operations this requires that you defeat an intelligence organization one of whose tasks is to deceive you. The implications are unpleasant and inadequate study has been made in this area to make definitive decisions.

There is a general method of deception against computer systems being used to launch fully automated attacks against other computer systems. The general method is to analyze the attacking system (the target) in terms of its use of responses from the defender and create sequences of responses that emulate the desired responses to the target. Because all such mechanisms published or widely used today are quite finite and relatively simplistic, with substantial knowledge of the attack mechanism, it is relatively easy to create a low-quality deception that will be effective. It is noteworthy, for example, that the Deception ToolKit [Coh99b], which was made publicly available in source form in 1998, is still almost completely effective against automated intelligence tools attempting to detect vulnerabilities. It seems that the widely used attack tools are not yet being designed to detect and counter deception.

That is not to say that red teams and intelligence agencies are not beginning to start to look at this issue. For example, in private conversations with defenders against select elite red teams the question often comes up of how to defeat the attackers when they undergo a substantial intelligence effort directed at defeating their attempts at deceptive defense. The answer is to increase the fidelity of the deception. This has associated costs, but as the attack tools designed to counter deception improve, so will the requirement for higher fidelity in deceptions.

2.6.10 Deception Mechanisms for Information Systems

The contents of Tables 2.2 and 2.3 are extracted from a previous paper on attack mechanisms [CPS+99]. It is intended to summarize the attack mechanisms that are viable deception techniques against information systems - in the sense that they induce or inhibit cognition at some level. All of the attack techniques in the original paper may be used as parts of overall deception processes, but only these are specifically useful as deception methods and specifically oriented toward information technology as opposed to the people that use and control these systems. We have explicitly excluded mechanisms used for observation only and included examples of how these techniques affect cognition and thus assist in deception and added information about deception levels in the target system.

2.7 Models of Deception of More Complex Systems

Larger cognitive systems can me modeled as being built up from smaller cognitive subsystems through some composition mechanism. Using these combined models we may analyze and create larger scale deceptions. To date there is no really good theory of composition for these sorts of systems and attempts to build theories of composition for security properties of even relatively simple computer networks have proven rather difficult. We can also take a top-down approach, but without the ability to link top-level objectives to bottom-level capabilities and without metrics for comparing alternatives, the problem space grows rapidly and results cannot be meaningfully compared.

2.7.1 Human Organizations

Humans operating in organizations and groups of all sorts have been extensively studied, but deception results in this field are quite limited. The work of Karrass [Kar70] (described earlier) deals with issues of negotiations involving small groups of people, but is not extended beyond that point. Military intelligence failures make good examples of organizational deceptions in which one organization attempts to deceive another. Hughes-Wilson describes failures in collection, fusion, analysis, interpretation, reporting, and listening to what intelligence is saying as the prime causes of intelligence blunders, and at the same time indicates that generating these conditions generally involved imperfect organizationally-oriented deceptions by the enemy [HW99]. John Keegan details a lot of the history of warfare and along the way described many of the deceptions that resulted in tactical advantage [Kee93]. Dunnigan and Nofi detail many examples of deception in warfare and, in some cases, detail how deceptions have affected organizations [DN95]. Strategic military deceptions have been carried out for a long time, but the theory of how the operations of groups lead to deception has never really been worked out. What we seem to have, from the time of Sun Tzu [Tzu83] to the modern day [DH82], is sets of rules that have withstood the test of time. Statements like *"It is far easier to lead a target astray by reinforcing the target's existing beliefs"* [57, p42] are stated and restated without deeper understanding, without any way to measure the limits of its effectiveness, and without a way to determine what beliefs an organization has. It sometimes seems we have

Mechanism	Levels
Cable cuts	HW
Fire	HW
Flood	HW
Earth movement	HW
Environmental control loss	HW
System maintenance	All
Trojan horses	All
Fictitious people	All
Resource availability manipulation	HW, OS
Spoofing and masquerading	All
Infrastructure interference	HW
Insertion in transit	All
Modification in transit	All
Sympathetic vibration	All
Cascade failures	All
Invalid values on calls	OS and up
Undocumented or unknown function exploitation	All
Excess privilege exploitation	App, Driver
Environment corruption	All
Device access exploitation	HW, Driver
Modeling mismatches	App and up
Simultaneous access exploitations	All
Implied trust exploitation	All
Interrupt sequence mishandling	Driver, OS
Emergency procedure exploitation	All
Desychronization and time-based attacks	All
Imperfect daemon exploits	Lib, App
Multiple error inducement	All
Viruses	All
Data diddling	OS and up
Electronic interference	HW
Repair-replace-remove information	All
Wire closet attacks	HW
Process bypassing	All
Content-based attacks	Lib and up
Restoration process corruption or misuse	Lib and up
Hangup hooking	HW, Lib, Driver, OS
Call forwarding fakery	HW
Input overflow	All
Illegal value insertion	All
Privileged program misuse	App, OS, Driver
Error-induced misoperation	All
Audit suppression	All
Induced stress failures	All
False updates	All
Network service and protocol attacks	HW, Driver, Proto
Distributed coordinated attacks	All
Man-in-the-middle	HW, Proto

Table 2.2: Deception Mechanisms and Levels (part 1)

Mechanism	Levels
Replay attacks	Proto, App, and up
Error insertion and analysis	All
Reflexive control	All
Dependency analysis and exploitation	All
Interprocess communication attacks	OS, Lib, Proto, App
Below-threshold attacks	All
Peer relationship exploitation	Proto, App, and up
Piggybacking	All
Collaborative misuse	All
Race conditions	All
Kiting	App and up
Salami attacks	App and up
Repudiation	App and up

Table 2.3: Deception Mechanisms and Levels (part 2)

not made substantial progress from when Sun Tzu originally told us that "All warfare is based on deception.

The systematic study of group deception has been under way for some time. In 1841, Mackay released his still famous and widely read book titled "Extraordinary Popular Delusions and the Madness of Crowds" [Mac89] in which he gives detailed accounts of the history of the largest scale deceptions and financial 'bubbles' of history to that time. It is astounding how relevant this is to modern times. For example, the recent bubble in the stock market related to the emergence of the Internet is incredibly similar to historical bubbles, as are the aftermaths of all of these events. The self-sustaining unwarranted optimism, the self fulfilling prophecies, the participation even by the skeptics, the exit of the originators, and the eventual bursting of the bubble to the detriment of the general public, all seem to operate even though the participants are well aware of the nature of the situation. While Mackay offers no detailed psychological accounting of the underlying mechanisms, he clearly describes the patterns of behavior in crowds that lead to this sort of group insanity.

Charles Handy [Han93] (Figure 2.5) describes how power and influence work in organizations. This leads to methods by which people with different sorts of power create changes in the overall organizational perspective and decision process. In deceptions of organizations, models of who stands where on which issues and methods to move them are vital to determining who to influence and in what manner in order to get the organization to move.

These principles have been applied without rigor and with substantial success for a long time.

Example: In World War II Germany, Hitler was the target of many of the allied strategic deceptions because the German organs of state were designed to grant him unlimited power. It didn't matter that Romel believed that the allies would attack at Normandy because Hitler was convinced that they would strike at Pas de Calais. All dictatorial regimes tend to be swayed by influencing the mind of a single key decision maker. At the same time we should not make the mistake of believing that this works at a tactical level. The German military in World War II was highly skilled at local decision making and field commanders were trained to innovate and take command when in command.

Military hierarchies tend to operate this way to a point, however, most military Juntas have a group decision process that significantly complicates this issue. For example, the FARC in Colombia have councils that make group decisions and cannot be swayed by convincing a single authority figure. Swaying the United States is very a complex process, while swaying Iraq is considerably easier, at least from a standpoint of identifying the target of deceptions. The previously cited works on individual human deception certainly provide us with the requisite rational for explaining

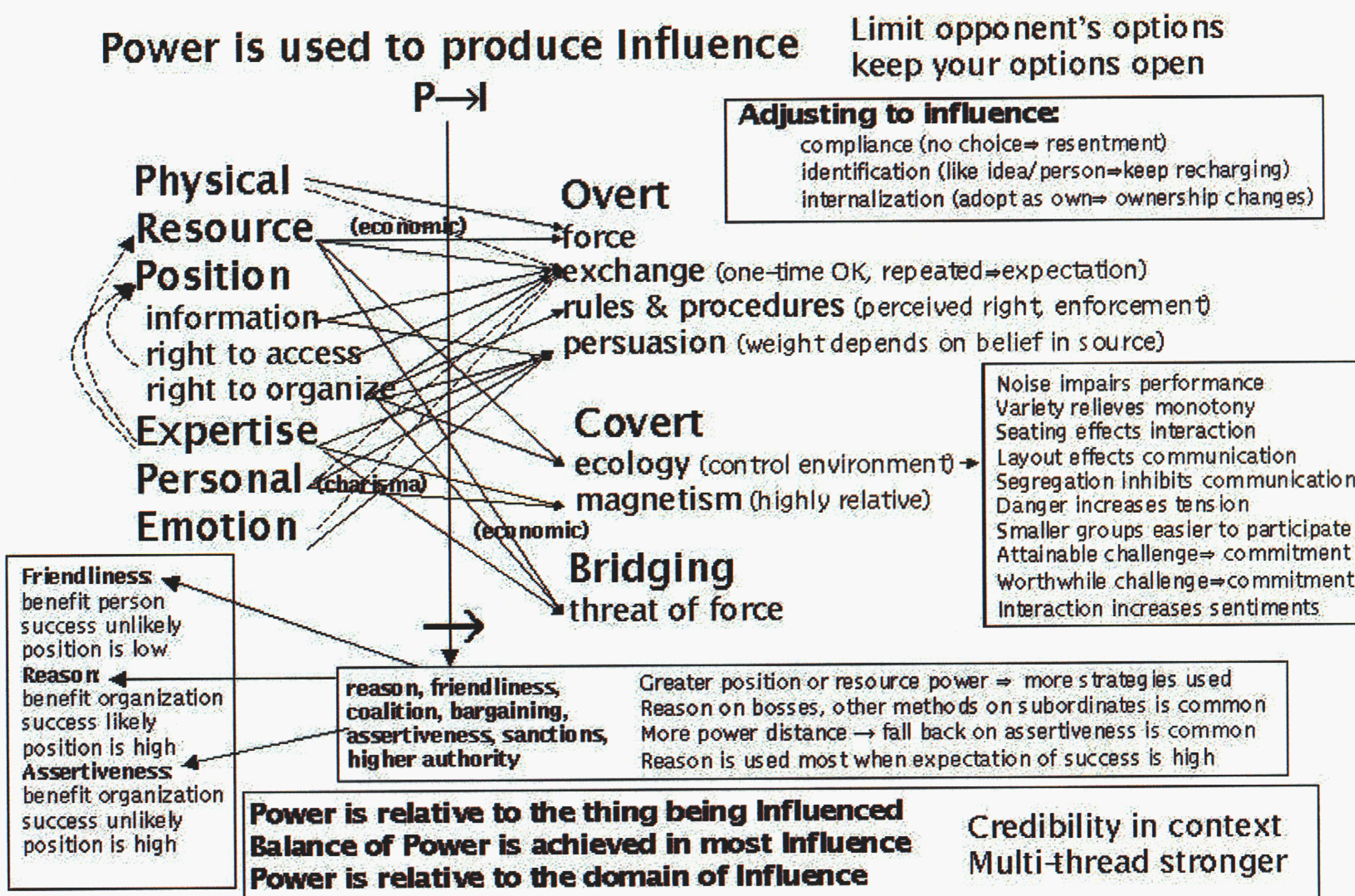


Figure 2.5: Power and Influence in Human Organizations

individual tendencies and the creation of conditions that tend to induce more advantageous behaviors in select circumstances, but how this translates into groups is a somewhat different issue.

Organizations have many different structures, but those who study the issue [Han93] have identified 4 classes of organizational structure that are most often encountered and which have specific power and influence associations: hierarchy, star, matrix, and network. In hierarchies orders come from above and reporting is done from lower level to higher level in steps. Going "over a supervisor's head" is considered bad form and is usually punished. These sorts of organizations tend to be driven by top level views and it is hard to influence substantial action except at the highest levels. In a star system all personnel report to a single central point. In small organizations this works well, but the center tends to be easily overloaded as the organization grows or as more and more information is fed into it. Matrix organizations tend to cause all of the individuals to have to serve more than one master (or at least manager). In these cases there is some redundancy, but the risk of inconsistent messages from above and selective information below exists. In a network organization, people form cliques and there is a tendency for information not to get everywhere it might be helpful to have it. Each organizational type has it features and advantages, and each has different deception susceptibility characteristics resulting from these structural features. Many organizations have mixes of these structures within them.

Deceptions within a group typically include; (1) members deceive other members, (2) members deceive themselves (e.g., "group think"), and (3) leader deceives members. Deception between groups typically include (1) leader deceives leader and (2) leader deceives own group members. Self deception applies to the individual acting alone.

> *Example: "group think", in which the whole organization may be mislead due to group processes/social norms. Many members of the German population in World War II became murderous even though under normal circumstances they never would have done the things they did.*

Complex organizations require more complex plans for altering decision processes. An effective deception against a typical government or large corporation may involve understanding a lot about organizational dynamics and happens in parallel with other forces that are also trying to sway the decision process in other directions. In such situations, the movement of key decision makers in specific ways tends to be critical to success, and this in turn depends on gaining access to their observables and achieving focus or lack of focus when and where appropriate. This can then lead to the need to gain access to those who communicate with these decision makers, their sources, and so forth.

> *Example: In the roll-up to the Faulkland Islands war between Argentina and the United Kingdom, the British were deceived into ignoring signs of the upcoming conflict by ignoring the few signs they say, structuring their intelligence mechanisms so as to focus on things the Argentines could control, and believing the Argentine diplomats who were intentionally asserting that negotiations were continuing when they were not. In this example, the Argentines had control over enough of the relevant sensory inputs to the British intelligence operations so that group-think was induced.*

Many studies have shown that optimal group sizes for small tightly knit groups tend to be in the range of 4-7 people. For tactical situations, this is the typical human group size. Whether the group is running a command center, a tank, or a computer attack team, smaller groups tend to lack cohesion and adequate skills, while larger groups become harder to manage in tight situations. It would seem that for tactical purposes, deceptions would be more effective if they could be successful at targeting a group of this size. Groups of this sort also have a tendency to have specialties with cross limited training. For example, in a computer attack group, a different individual will likely be an expert on one operating system as opposed to another. A hardware expert, a fast systems programmer / administrator, appropriate operating system and other domain experts, an information fusion person, and a skilled Internet collector may emerge. No systematic testing of these notions has been done to date but personal experience shows it to be true. Recent work in large group collaboration using information technology to augment normal human capabilities have show limited promise. Experiments will be required to determine whether this is an effective tool in carrying out or defeating deceptions, as well as how such a tool can be exploited so as to deceive its users.

The National Research Council [NRC98] discusses models of human and organizational behavior and how automation has been applied in the modeling of military decision making. This includes a wide range of computer-based modeling systems that have been developed for specific applications and is particularly focused on military and combat situations. Some of these models would appear to be useful in creating effective models for simulation of behavior under deceptions and several of these models are specifically designed to deal with psychological factors. This field is still very new and the progress to date is not adequate to provide coverage for analysis of deceptions, however, the existence of these models and their utility for understanding military organizational situations may be a good foundation for further work in this area.

2.7.2 Computer Network Deceptions

Computer network deceptions essentially never exist without people involved. The closest thing we see to purely computer to computer deceptions have been feedback mechanisms that induce livelocks or other denial of service impacts. These are the result of misinformation passing between computers.

> *Examples include the electrical cascade failures in the U.S. power grid, [WSC] telephone system cascade failures causing widespread long distance service outages, [Pek90] and inter-system cascades such as power failures bringing down telephone switches required to bring power stations back up. [Pek90]*

But the notion of deception, as we define it, involves intent, and we tend to attribute intent only to human actors at this time. There are, of course, programs that display goal directed behavior, and we will not debate the issue further except to indicate that, to date, this has not been used for the purpose of creating network deceptions without human involvement.

Individuals have used deception on the Internet since before it became the Internet. In the Internet's predecessor, the ARPAnet, there were some rudimentary examples of email forgeries in which email was sent under an alias - typically as a joke. As the Internet formed and become more widespread, these deceptions continued in increasing numbers and with increasing variety. Today, person to person and person to group deception in the Internet is commonplace and very widely practiced as part of the notion of anonymity that has pervaded this media. Some examples of papers in this area include:

> *"Gender Swapping on the Internet" [Van] was one of the original "you can be anyone on the Internet" descriptions. It dealt with players in MUDs (Multi-User Dungeon), which are multiple-participant virtual reality domains. Players soon realized that they could have multiple online personalities, with different genders, ages, and physical descriptions. The mind behind the keyboard often chooses to stay anonymous, and without violating system rules or criminal laws, it is difficult or impossible for ordinary players to learn many real-world identities.*
>
> *"Cybernetic Fantasies: Extended Selfhood in a Virtual Community" by Mimi Ito, from 1993 [Ito93], is a first-person description of a Multi-User Dungeon (MUD) called Farside, which was developed at a university in England. By 1993 it had 250 players. Some of the people using Farside had characters they maintained in 20 different virtual reality*

MUDs. Ito discusses previous papers, in which some people went to unusual lengths such as photos of someone else, to convince others of a different physical identity.

"Dissertation: A Chatroom Ethnography" by Mark Peace [Pea00], is a more recent study of Internet Relay Chat (IRC), a very popular form of keyboard to keyboard communication. This is frequently referred to as Computer Mediated Communication (CMC). Describing first-person experiences and observation, Peace believes that many users of IRC do not use false personalities and descriptions most of the time. He also provides evidence that IRC users do use alternate identities.

Daniel Chandler writes, "In a 1996 survey in the USA, 91% of homepage authors felt that they presented themselves accurately on their web pages (though only 78% believed that other people presented themselves accurately on their home pages!)" [Cha01]

Criminals have moved to the Internet environment in large numbers and use deception as a fundamental part of their efforts to commit crimes and conceal their identities from law enforcement. While the specific examples are too numerous to list, there are some common threads, among them that the same criminal activities that have historically worked person to person are being carried out over the Internet with great success.

- Identity theft is one of the more common deceptions based on attacking computers. In this case, computers are mined for data regarding an individual and that individual's identity is taken over by the criminal who then commits crimes under the assumed name. The innocent victim of the identity theft is often blamed for the crimes until they prove themselves innocent.
- One of the most common Internet-based deceptions is an old deception of sending a copier supply bill to a corporate victim. In many cases the internal controls are inadequate to differentiate a legitimate bill from a fraud and the criminal gets paid illegitimately.
- Child exploitation is commonly carried out by creating friends under the fiction of being the same age and sex as the victim. Typically a 40 year old pedophile will engage a child and entice them into a meeting outside the home. In some cases there have been resulting kidnappings, rapes, and even murders.

During the cyber conflict between the Palestinian Liberation Organization (PLO) and a group of Israeli citizens that started early in 2001, one PLO cyber terrorist lured an Israeli teenager into a meeting and kidnapped and killed the teen. In this case the deception was the simulation of a new friend made over the Internet:

> *The Internet "war" assumed new dimensions here last week, when a 23-year-old Palestinian woman, posing as an American tourist, apparently used the Internet to lure a 16-year-old Israeli boy to the Palestinian Authority areas so he could be murdered. -* Hanan Sher, The Jerusalem Report, 2001/02/10

Larger scale deceptions have also been carried out over the Internet. For example, one of the common methods is to engage a set of 'shills' who make different points toward the same goal in a given forum. While the forum is generally promoted as being even handed and fair, the reality is that anyone who says something negative about a particular product or competitor will get lambasted. This has the social effect of causing distrust of the dissenter and furthering the goals of the product maker. The deception is that the seemingly independent members are really part of the same team, or in some cases, the same person. In another example, a student at a California university made false postings to a financial forum that drove down the price of a stock that the student had invested in derivatives of. The net effect was a multi-million dollar profit for the student and the near collapse of the stock.

The largest scale computer deceptions tend to be the result of computer viruses. Like the mass hysteria of a financial bubble, computer viruses can cause entire networks of computers to act as a rampaging group. It turns out that the most successful viruses today use human behavioral characteristics to induce the operator to foolishly run the virus which, on its own, could not reproduce. They typically send an email with an infected program as an attachment. If the infected program is run it then sends itself in email to other users this user communicates with, and so forth. The deception is the method that convinces the user to run the infected program. To do this, the program might be given an enticing name, or the message may seem like it was really from a friend asking the user to look at something, or perhaps the program is simply masked so as to simulate a normal document.

> *In one case a computer virus was programmed to silently dial out on the user's phone line to a telephone number that generated revenues to the originator of the virus (a 900 number). This example shows how a computer system can be attacked while the user is completely unaware of the activity.*

These are deceptions that act across computer networks against individuals who are attached to the network. They are targeted at the millions of individuals who might receive them and, through the viral mechanism, distribute the financial burden across all of those individuals. They are a form of a "Salami" attack in which small amounts are taken from many places with large total effect.

2.7.3 Implications

These examples would tend to lead us to believe that effective defensive deceptions against combinations of humans and computers are easily carried out to substantial effect, and indeed that appears to be true, if the only objective is to fool a casual attacker in the process of breaking into a system from outside or escalating privilege once they have broken in. For other threat profiles, however, such simplistic methods will not likely be successful, and certainly not remain so for long once they are in widespread use. Indeed, all of these deceptions have been oriented only toward being able to observe and defend against attackers in the most direct fashion and not oriented toward the support of larger deceptions such as those required for military applications.

There have been some studies of interactions between people and computers. Some of the typical results include the notions that people tend to believe things the computers tell them, humans interacting through computers tend to level differences of stature, position, and title, that computer systems tend to trust information from other computer systems excessively, that experienced users to interact differently than less experienced ones, the ease of lying about identities and characteristics as demonstrated by numerous stalking cases, and the rapid spread viruses as an interaction between systems with immunity to viruses (by people) for limited time periods. The Tactical Decision Making Under Stress (TADMUS) program is an example of a system designed to mitigate decision errors caused by cognitive overload, which have been documented through research and experimentation [SSC].

Sophisticated attack groups tend to be small, on the order of 4-7 people in one room, or operate as a distributed group perhaps as many as 20 people can loosely participate. Most of the most effective groups have apparently been small cells of 4 to 7 people or individuals with loose connections to larger groups. Based on activities seen to date, but without a comprehensive study to back these notions up, less than a hundred such groups appear to be operating overtly today, and perhaps a thousand total groups would be a good estimate based on the total activities detected in openly available information. A more accurate evaluation would require additional research, specifically including the collection of data from substantial sources, evaluation of operator and group characteristics (e.g., times of day, preferred targets, typing characteristics, etc.), and tracking of modus operandi of perpetrators. In order to do this, it would be prudent to start to create sample attack teams and do substantial experiments to understand the internal development of these team, team characteristics over time, team makeup, develop capabilities to detect and differentiate teams, and test out these capabilities in a larger environment. Similarly, the ability to reliably deceive these groups will depend largely on gaining understanding about how they operate.

We believe that large organizations are only deceived by strategic application of deceptions against individuals and small groups. While we have no specific evidence to support this, ultimately is must be true to some extent because groups don't make decisions without individuals making decisions. While there may be different motives for different individuals and groups insanity of a sort may be part of the overall effect, there nevertheless must be specific individuals and small groups that are deceived in order for them to begin to convey the overall message to other groups and individuals. Even in the large-scale perception management campaigns involving massive efforts at propaganda, individual opinions are affected first, small groups follow, and then larger groups become compliant under social pressures and belief mechanisms.

Thus the necessary goal of creating deceptions is to deceive individuals and then small groups that those individuals are part of. This will be true until targets develop far larger scale collaboration capabilities that might allow them to make decisions on a different basis or change the cognitive structures of the group as a whole. This sort of technology is not available at present in a manner that would reduce effectiveness of deception and it may never become available.

Clearly, as deceptions become more complex and the systems they deal with include more and more diverse components, the task of detailing deceptions and their cognitive nature becomes more complex. It appears that there is regular structure in most deceptions involving large numbers of systems of systems because the designers of current widespread attack deceptions have limited resources. In such cases it appears that a relatively small number of factors can serve to model the deceptive elements, however, large scale group deception effects may be far more complex to understand and analyze because of the large number of possible interactions and complex sets of interdependences involved in cascade failures and similar phenomena. If deception technology continues to expand and analytical and implementation capabilities become more substantial, there is a tremendous potential for highly complex deceptions wherein many different systems are involved in highly complex and irregular interactions. In such an environment, manual analysis will not be capable of dealing with the issues and automation will be required in order to both design the deceptions and counter them.

2.7.4 Experiments and the Need for an Experimental Basis

One of the more difficult things to accomplish in this area is meaningful experiments. While a few authors have published experimental results in information protection, far fewer have attempted to use meaningful social science methodologies in these experiments or to provide enough testing to understand real situations. This may be because of the difficulty and high cost of each such experiment and the lack of funding and motivation for such efforts. We have identified this as a critical need for future work in this area.

If one thing is clear from our efforts it is the fact that too few experiments have been done to understand how deception works in defense of computer systems and, more generally, too few controlled experiments have been done to understand the computer attack and defense processes and to characterize them. Without a better empirical basis, it will be hard to make scientific conclusions about such efforts.

While anecdotal data can be used to produce many interesting statistics, the scientific utility of those statistics is very limited because they tend to reflect only those examples that people thought worthy of calling out. We get only *"lies, damned lies, and statistics."*

Experiments to Date From the time of the first published results on honeypots, the total number of published experiments performed in this area appear to be very limited. While there have been hundreds of published experiments by scores of authors in the area of human deception, articles on computer deception experiments can be counted on one hand.

Cohen provided a few examples of real world effects of deception [Coh99b], but performed no scientific studies of the effects of deception on test subjects. While he did provide a mathematical analysis of the statistics of deception in a networked environment, there was no empirical data to confirm or refute these results [Coh99a].

The HoneyNet Project [Hon] is a substantial effort aimed at placing deception system in the open environment for detection and tracking of attack techniques. As such, they have been largely effective at luring attackers. These lures are real systems placed on the Internet with the purpose of being attacked so that attack methods can be tracked and assessed. As deceptions, the only thing deceptive about them is that they are being watched more closely than would otherwise be apparent and known faults are intentionally not being fixed to allow attacks to proceed. These are highly effective at allowing attackers to enter because they are extremely high fidelity, but only for the purpose they are intended to provide. They do not, for example, include any user behaviors or content of interest. They are quite effective at creating sites that can be exploited for attack of other sites. For all of the potential benefit, however, the HoneyNet project has not performed any controlled experiments to understand the issues of deception effectiveness.

Red teaming (i.e., finding vulnerabilities at the request of defenders) [Coh98b] has been performed by many groups for quite some time. The advantage of red teaming is that it provides a relatively realistic example of an attempted attack. The disadvantage is that it tends to be somewhat artificial and reflective of only a single run at the problem. Real systems get attacked over time by a wide range of attackers with different skill sets and approaches. While many red teaming exercises have been performed, these tend not to provide the scientific data desired in the area of defensive deceptions because they have not historically been oriented toward this sort of defense.

Similarly, war games played out by armed services tend to ignore issues of information system attacks because the exercises are quite expensive and by successfully attacking information systems that comprise command and control capabilities, many of the other purposes of these war games are defeated. While many recognize that the need to realistically portray effects is important, we could say the same thing about nuclear weapons, but that doesn't justify dropping them on our forces for the practice value.

The most definitive experiments to date that we were able to find on the effectiveness of low-quality computer deceptions against high quality computer assisted human attackers were performed by RAND [GWM+00]. Their experiments with fairly generic deceptions operated against high quality intelligence agency attackers demonstrated substantial effectiveness for short periods of time. This implies that under certain conditions (i.e., short time frames, high tension, no predisposition to consider deceptions, etc.) these deceptions may be effective.

The total number of controlled experiments to date involving deception in computer networks appear to be less than 20, and the number involving the use of deceptions for defense are limited to the 10 or so from the RAND study. Clearly this is not enough to gain much in the way of knowledge and, just as clearly, many more experiments are required in order to gain a sound understanding of the issues underlying deception for defense.

Experiments We Believe Are Needed At This Time In this study, a large set of parameters of interest have been identified and several hypotheses put forth. We have some anecdotal data at some level of detail, but we don't have a set of scientific data to provide useful metrics for producing scientific results. In order for our models to be effective in producing increased surety in a predictive sense we need to have more accurate information.

The clear solution to this dilemma is the creation of a set of experiments in which we use social science methodologies to create, run, and evaluate a substantial set of parameters that provide us with better understanding and specific metrics and accuracy results in this area. In order for this to be effective, we must not only create defenses, but also come to understand how attackers work and think. For this reason, we will need to create red teaming experiments in which we study both the attackers and the effects of defenses on the attackers. In addition, in order to isolate the effects of deception, we need to create control groups, and experiments with double blinded data collection.

2.8 Analysis and Design of Deceptions

A good model should be able to explain, but a good scientific model should be able to predict and a good model for our purposes should help us design as well. At a minimum, the ability to predict leads to the ability to design by random variation and selective survival with the survival evaluation being made based on prediction. In most cases, it is a lot more efficient to have the ability to create design rules that are reflective of some underlying structure.

Any model we build that is to have utility must be computationally reasonable relative to the task at hand. Far more computation is likely to be available for a large-scale strategic deception than for a momentary tactical deception, so it would be nice to have a model that scales well in this sense. Computational power is increasing with time, but not at such a rate that we will ever be able to completely ignore computational complexity in problems such as this.

A fundamental design problem in deception lies in the fact that deceptions are generally thought of in terms of presenting a desired story to the target, while the available techniques are based on what has been found to work. In other words, there is a mismatch between available deception techniques and technologies and objectives.

2.8.1 A Language for Analysis and Design of Deceptions

Rather than focus on what we wish to do, our approach is to focus on what we can do and build up 'deception programs' from there. In essence, our framework starts with a programming language for human deception by finding a set of existing primitives and creating a syntax and semantics for applying these primitives to targets. We can then associate metrics with the elements of the programming language and analyze or create deceptions that optimize against those metrics.

The framework for human deception then has three parts:

- **A set of primitive techniques:** The set of primitive techniques is extensive and is described hierarchically based on the model shown above, with each technique associated with one or more of Observables, Actions, Assessments, Capabilities, Expectations, and Intent and causing an effect on the situation depicted by the model.

- **Properties of those techniques:** Properties of techniques are multi-dimensional and include all of the properties discussed in this report. This includes, but is not limited to, resources consumed, effect on focus of attention, concealment, simulation, memory requirements and impacts, novelty to target, certainty of effect, extent of effect, timeliness of effect, duration of effect, security requirements, target system resource limits, deceiver system resource limits, the effects of small changes, organizational structure, knowledge, and constraints, target knowledge requirements, dependency on predisposition, extent of change in target mind set, feedback potential and availability, legality, unintended consequences, the limits of modeling, counterdeception, recursive properties, and the story to be told. These are the same properties of deception discussed under "The Nature Of Deception" earlier.

- **A syntax and semantics for applying and optimizing the properties:** This is a language that has not yet been developed for describing, designing, and analyzing deceptions. It is hoped that this language and the underlying database and simulation mechanism will be developed in subsequent efforts.

The astute reader will recognize this as the basis for a computer language, but it has some differences from most other languages, most fundamentally in that it is probabilistic in nature.

Deception Property	Technique 1	...	Technique n
name	Audit Suppression		
general concept	packet flooding of audit mechanisms		
means	using a distributed set of intermediaries		
target type	computer		
resources consumed	reveals intermediaries which will be disabled with time		
effect on focus of attention	induces focus on this attack		
concealment	conceals other actions from target audit and analysis		
simulation	n/a		
memory requirements and impacts	overruns target memory capacity		
novelty to target	none - they have seen similar things before		
ncertainty of effect	80% effective if intel is right		
extent of effect	reduces audits by 90% if effective		
timeliness of effect	takes 30 seconds to start		
duration of effect	until ended or intermediaries are disabled		
security requirements	must conceal launch points and intermediaries		
target system resource limits	memory capacity, disk storage, CPU time		
deceiver system resource limits	number of intermediaries for attack, prepositioned assets lost with attack		
the effects of small changes	nonlinear effect on target with break point at effectiveness threshold		
organizational structure and constraints	Going after known main audit server which will impact whole organization audits		
target knowledge	OS type and release		
dependency on predisposition	Must be proper OS type and release to work		
extent of change in target mind set	Large change - it will interrupt them - they will know they are being attacked		
feedback potential and availability	Feedback apparent in response behavior observed against intermediaries and in other fora		
legality	Illegal except at high intensity conflict - possible act of war		
unintended consequences	Impacts other network elements, may interrupt other information operations, may result in increased target security		
the limits of modeling	Unable to model overall network effects		
counterdeception	If feedback known or attack anticipated, easy to deceive attacker		
recursive properties	only through counter deception		
possible deception story	We are concealing something - they know this - but they don't know what		

Table 2.4: Deception Properties and Techniques

	Table 7.1 from "Emerging Viruses"	Computer Viruses	Manual Attacks
1	Stability in environment	Stability in environment	Stability in environment
2	Entry into host - portal of entry	Entry into host - portal of entry	Entry into host - portal of entry
3	Localization in cells near portal of entry	Localization in software near portal of entry	Localization near portal of entry
4	Primary replication	Primary replication	Primary modifications
5	Non-specific immune response	Non-specific immune response	Non-specific immune response
6	Spread from primary site (blood, Nerves)	Spread from primary site (disk, comms)	Spread from primary site (privilege expansion)
7	Cells and tissue tropism	Program and data tropism	Program and data tropism (hiding)
8	Secondary replication	Secondary replication	Secondary replication
9	Antibody and cellular immune response	Human and program immune response	Human and program immune response
10	Release from host	Release from host	Release from host (spread on)

Table 2.5: Pathogenesis of Attacks

While most programming languages guarantee that when you combine two operators together in a sequence you get the effect of the first followed by the effect of the second, in the language of deception, a sequence of operators produces a set of probabilistic changes in perceptions of all parties across the multi-dimensional space of the properties of deception. It will likely be effective to "program" in terms of desired changes in deception properties and allow the computer to "compile" those desired changes into possible sequences of operators. The programming begins with a 'firing table' of some sort that looks something like Table 2.4, but with many more columns filled in and many more details under each of the rows. Partial entries are provided for technique 1 which, for this example, we will choose as 'audit suppression' by packet flooding of audit mechanisms using a distributed set of previously targeted intermediaries.

Considering that the total number of techniques is likely to be on the order of several hundred and the vast majority of these techniques have not not been experimentally studied, the level of effort required to build such a table and make it useful will be considerable.

2.8.2 Attacker Strategies and Expectations

For a moment, we will pause from the general issue of deception and examine more closely the situation of an attacker attempting to exploit a defender through information system attack. In this case, there is a commonly used attack methodology that subsumes other common methodologies and there are only three known successful attack strategies identified by simulation and verified against empirical data. We start with some background.

The pathogenesis of diseases has been used to model the process of breaking onto computers and it offers an interesting perspective [Coh00c]. In this view, the characteristics of an attack are given in terms of the survival of the attack method. Table 2.5 shows the pathogenesis of a biological virus attack from [Coh00c] compared to the the analogous pathogenesis for computer viruses and for manual attacks on a computer system.

This particular perspective on attack as a biological process ignores one important facet of the problem, and that is the preparation process for an intentional and directed attack. In the case of most computer viruses, targeting is not an issue. In the case of an intelligent attacker, there is generally a set of capabilities and an intent behind the attack. Furthermore, survival (stability in the environment) would lead us to the conclusion that a successful attacker who does not wish to be traced back to their origin will use an intelligence process including personal risk reduction as part of their overall approach to attack. This in turn leads to an intelligence process that precedes the actual attack.

The typical attack methodology consists of:

1. intelligence gathering, securing attack infrastructure, tool development, and other preparations,
2. system entry (beyond default remote access),
3. privilege expansion,
4. subversion, typically involving planting capabilities and verifying over time, and
5. exploitation.

There are loops from higher numbers to lower numbers so that, for example, privilege expansion can lead back to intelligence and system entry or forward to subversion, and so forth. In addition, attackers have expectations throughout this process that adapt based on what has been seen before this attack and within this attack. Clean up, observation of effects, and analysis of feedback for improvement are also used throughout the attack process.

Extensive simulation has been done to understand the characteristics of successful attacks and defenses [Coh99c]. Among the major results of this study were a set of successful strategies for attacking computer systems. It is particularly interesting that these strategies are similar to classic military strategies because the simulation methods used were not designed from a strategic viewpoint, but were based solely on the mechanisms in use and the times, detection, reaction, and other characteristics associated with the mechanisms themselves. Thus the strategic information that fell out of this study was not biased by its design but rather emerged as a result of the metrics associated with different techniques. The successful attack strategies identified by this study included:

1. speed,
2. stealth, and
3. overwhelming force.

Slow, loud attacks tend to be detected and reacted to fairly easily. A successful attacker can use combinations of these in different parts of an attack. For example, speed can be used for a network scan, stealth for system entry, speed for privilege expansion and planting of capabilities, stealth for verifying capabilities over time, and overwhelming force for exploitation. This is a typical pattern today.

Substantial red teaming and security audit experience has led to some speculations that follow the general notions of previous work on individual deception. It seems clear from experience that people who use computers in attacks:

1. tend to trust what the computers tell them unless it is far outside normal expectations,
2. use the computer to automate manual processes and not to augment human reasoning, and
3. tend to have expectations based on prior experience with their tools and targets.

If this turns out to be true, it has substantial implications for both attack and defense. Experiments should be undertaken to examine these assertions as well as to study the combined deception properties of small groups of people working with computers in attacking other systems. Unfortunately, current data is not adequate to thoroughly understand these issues. There may be other strategies developed by attackers, other attack processes undertaken, and other tendencies that have more influence on the process. We will not know this until extensive experimentation is done in this area.

2.8.3 Defender Strategies and Expectations

From the deceptive defender's perspective, there also seem to be a limited set of strategies.

- **Computer Only:** If the computer is being used for a fully automated attack, analysis of the attack tool or relatively simply automated response mechanisms are highly effective at maintaining the computer's expectations, dazzling the computer to induce unanticipated processing and results, feeding false information to the computer, or in some cases, causing the computer to crash. We have been able to easily induce or suppress signal returns to an attacking computer and have them seen as completely credible almost no matter how ridiculous they are. Whether this will continue and to what extent it will continue in the presence of a sophisticated hostile environment remain to be seen.

- **People Only:** Manual attack is very inefficient so it is rarely used except in cases where very specific targets are involved. Because humans do tend to see what they expect to see, it is relatively easy to create high fidelity deceptions by redirecting traffic to a honey pot or other such system. Indeed, this transition can even be made fairly early in an attack without most human attackers noticing it. In this case there are three things we might want to do:

  1. maintain the attackers expectations to consume their time and effort,
  2. slowly change their expectations to our advantage at a rate that is not noticeable by typical humans (e.g., slow the computer's response minute by minute till it is very slow and the attacker is wasting lots of time and resources), and
  3. create cognitive dissonance to force them to think more deeply about what is going on, wonder if they have been detected, and induce confusion in the attacker.

- **People With Poorly Integrated Computers:** This is the dominant form of efficient widespread attack today. In this form, people use automated tools combined with short bursts of human activity to carry out attacks.

  The intelligence process is almost entirely done by scanning tools which (1) can be easily deceived and (2) tend to be believed. Such deceptions will only be disbelieved if inconsistencies arise between tools, in which case the tools will initially be suspected.

  System entry is either automated with the intelligence capability or automated at a later time when the attacker notices that an intelligence sweep has indicated a potential vulnerability. Results of these tools will be believed unless they are incongruous with normal expectations.

  Privilege expansion is either fully automated or has a slight manual component to it. It typically involves the loading of a toolkit for the job followed by compilation and/or execution. This typically involves minimal manual effort. Results of this effort are believed unless they are incongruous with normal expectations.

  Planting capabilities is typically nearly automated or fully automated. Returning to verify over time is typically automated with time frames substantially larger than attack times. This will typically involve minimal manual effort. Results of this effort will be believed unless they are incongruous with normal expectations.

  Exploitation is typically done under one-shot or active control. A single packet may trigger a typical exploit, or in some cases the exploit is automatic and ongoing over an extended period of time. This depends on whether speed, stealth, or force is desired in the exploitation phase. This causes observables that can be validated by the attacker. If the observables are not present it might generate deeper investigation by the attacker. If there are plausible explanations that can be discovered by the attacker they will likely be believed.

- **People With Well Integrated Computers:** This has not been observed to date. People are not typically augmenting their intelligence but rather automating tasks with their computers.

As in the case with attacker strategies, few experiments have been undertaken to understand these issues in detail, but preliminary experiments appear to confirm these notions.

2.8.4 Planning Deceptions

Several authors have written simplistic analyses and provided rules of thumb for deception planning. There are also some notions about planning deceptions under the present model using the notions of low, middle, and high level cognition to differentiate actions and create our own rules of thumb with regard to our cognitive model. But while notions are fine for contemplation, scientific understanding in this area requires an experimental basis.

According to [Arm98] a 5-step process is used for military deception. (1) Situation analysis determines the current and projected enemy and friendly situation, develops target analysis, and anticipates a desired situation. (2) Deception objectives are formed by desired enemy action or non-action as it relates to the desired situation and friendly force objectives. (3) Desired [target] perceptions are developed as a means to generating enemy action or inaction based on what the enemy now perceives and would have to perceive in order to act or fail to act - as desired. (4) The information to be conveyed to or kept from the enemy is planned as a story or sequence, including the development and analysis of options. (5) A deception plan is created to convey the deception story to the enemy.

These steps are carried out by a combination of commander and command staff as an embedded part of military planning. Because of the nature of military operations, capabilities that are currently available and which have been used in training exercises and actual combat are selected for deceptions. This drives the need to create deception capabilities that are flexible enough to support the commander's needs for effective use of deceptions in a combat situation. From a standpoint of information technology deceptions, this would imply that, for example, a deceptive feint or movement of forces behind smoke screens with sonic simulations of movement should be supported by simulated information operations that would normally support such action and concealed information operations that would support the action being covered by the feint.

Deception maxims are provided to enhance planner understanding of the tools available and what is likely to work [Arm98]:

Magruder's principles - the exploitation of perceptions: It is easier to maintain an existing belief than to change it or create a new one.

Limitations of human information processing: The law of small numbers (once you see something twice it is taken as a maxim), and susceptibility to conditioning (the cumulative effect of small changes). These are also identified and described in greater detail in Gilovich [Gil91].

Cry-Wolf: This is a variant on susceptibility to conditioning in that, after a seeming threat appears again and again to be innocuous, it tends to be ignored and can be used to cover real threats.

Jones' Dilemma: Deception is harder when there are more information channels available to the target. On the other hand, the greater the number of 'controlled channels', the better it is for the deception.

A choice among deception types: In "A-type" deception, ambiguity is introduced to reduce the certainty of decisions or increase the number of available options. In "M-type" deception, misdirection is introduced to increase the victim's certainty that what they are looking for is their desired (deceptive) item.

Axelrod's contribution - the husbanding of assets: Some deceptions are too important to reveal through their use, but there is a tendency to over protect them and thus lose them by lack of application. Some deception assets become useless once revealed through use or overuse. In cases where strategic goals are greater than tactical needs, select deceptions should be held in reserve until they can be used with greatest effect.

A sequencing rule: Sequence deceptions so that the deception story is portrayed as real for as long as possible. The most clear indicators of deception should be held till the last possible moment. Similarly, riskier elements of a deception (in terms of the potential for harm if the deception is discovered) should be done later rather than earlier so that they may be called off if the deception is found to be a failure.

The importance of feedback: A scheme to ensure accurate feedback increases the chance of success in deception.

The Monkey's Paw: Deceptions may create subtle and undesirable side effects. Planners should be sensitive to such possibilities and, where prudent, take steps to minimize these effects.

Care in the designed and planned placement of deceptive material: Great care should be used in deceptions that leak notional information to targets. Apparent windfalls are subjected to close scrutiny and often disbelieved. Genuine leaks often occur under circumstances thought improbable.

Deception failures are typically associated with (1) detection by the target and (2) inadequate design or implementation. Many examples of this are given in [Arm98].

As a doctrinal matter, Battlefield deception involves the integration of intelligence support, integration and synchronization, and operations security.

Intelligence Support: Battlefield deceptions rely heavily on timely and accurate intelligence about the enemy. To make certain that deceptions are effective, we need to know (1) how the target's decision and intelligence cycles work, (2) what type of deceptive information they are likely to accept, (3) what source they rely on to get their intelligence, (4) what they need to confirm their information, and (5) what latitude they have in changing their operations. This requires both advanced information for planning and real-time information during operations.

Integration and Synchronization: Once we know the deception plan we need to synchronize it with the true combat operations for effect. History has shown that for the greatest chance of success, we need to have plans that are: (1) flexible, (2) doctrinally consistent with normal operations, (3) credible as to the current situation, and (4) simple enough to not get confused during the heat of battle. Battlefield deceptions almost always involve the commitment of real forces, assets, and personnel.

Operations Security: OPSEC is the defensive side of intelligence. In order for a deception to be effective, we must be able to deny access to the deceptive nature of the effort while also denying access to our real intentions. Real intentions must be concealed, manipulated, distorted, and falsified though OPSEC.

> *"OPSEC is not an administrative security program. OPSEC is used to influence enemy decisions by concealing specific, operationally significant information from his intelligence collection assets and decision processes. OPSEC is*

a concealment aspect for all deceptions, affecting both the plan and how it is executed" [Arm98]

In the DoD context, it must be assumed that any enemy is well versed in DoD doctrine. This means that anything too far from normal operations will be suspected of being a deception even if it is not. This points to the need to vary normal operations, keep deceptions within the bounds of normal operations, and exploit enemy misconceptions about doctrine. Successful deceptions are planned from the perspective of the targets.

The DoD has defined a set of factors in deceptions that should be seriously considered in planning. It is noteworthy that these rules are clearly applicable to situations with limited time frames and specific objectives and, as such, may not apply to situations in information protection where long-term protection or protection against nebulous threats are desired.

Policy: Deception is never an end in itself. It must support a mission.

Objective: A specific, realistic, clearly defined objective is an absolute necessity. All deception actions must contribute to the accomplishment of that objective.

Planning: Deception should be addressed in the commander's initial guidance to staff and the staff should be engaged in integrated deception and operations planning.

Coordination: The deception plan must be in close coordination with the operations plan.

Timing: Sufficient time must be allowed to: (1) complete the deception plan in an orderly manner, (2) effect necessary coordination, (3) promulgate tasks to involved units, (4) present the deception to the enemy decision-maker through their intelligence system, (5) permit the enemy decision maker to react in the desired manner, including the time required to pursue the desired course of action.

Security: Stringent security is mandatory. OPSEC is vital but must not prevent planning, coordination, and timing from working properly.

Realism: It must look realistic.

Flexibility: The ability to react rapidly to changes in the situation and to modify deceptive action is mandatory.

Intelligence: Deception must be based on the best estimates of enemy intelligence collection and decision-making processes and likely intentions and reactions.

Enemy Capabilities: The enemy commander must be able to execute the desired action.

Friendly Force Capabilities: Capabilities of friendly forces in the deception must match enemy estimates of capabilities and the deception must be carried out without unacceptable degradation in friendly capabilities.

Forces and Personnel: Real forces and personnel required to implement the deception plan must be provided. Notional forces must be realistically portrayed.

Means: Deception must be portrayed through all feasible and available means.

Supervision: Planning and execution must be continuously supervised by the deception leader. Actions must be coordinated with the objective and implemented at the proper time.

Liaison: Constant liaison must be maintained with other affected elements to assure that maximum effect is attained.

Feedback: A reliable method of feedback must exist to gage enemy reaction.

Deception of humans and automated systems involves interactions with their sensory capabilities. For people, this includes (1) visual (e.g., dummies and decoys, camouflage, smoke, people and things, and false vs. real sightings), (2) Olfactory (e.g., projection of odors associated with machines and people in their normal activities at that scale including toilet smells, cooking smells, oil and gas smells, and so forth), (3) sonic (e.g., directed against sounding gear and the human ear blended with real sounds from logical places and coordinated to meet the things being simulated at the right places and times) (4) electronic (i.e., manipulative electronic deception, simulative electronic deception, and imitative electronic deception).

Resources (e.g., time, devices, personnel, equipment, materiel) are always a consideration in deceptions as are the need to hide the real and portray the false. Specific techniques include (1) feints, (2) demonstrations, (3) ruses, (4) displays, (5) simulations, (6) disguises, and (7) portrayals [Arm98].

2.8.5 A Different View of Deception Planning Based on the Model from this Study

A typical deception is carried out by the creation and invocation of a deception plan. Such a plan is normally based on some set of reasonably attainable goals and time frames, some understanding of target characteristics, and some set of resources which are made available for use. It is the deception planner's objective to attain the goals with the provided resources within the proper time frames. In defending information systems through deception our objective is to deceive human attackers and defeat the purposes of the tools these humans develop to aid them in their attacks. For this reason, a framework for human deception is vital to such an undertaking.

All deception planning starts with the objective. It may work its way back toward the creation of conditions that will achieve that objective or use that objective to 'prune' the search space of possible deception methods. While it is tempting for designers to come up with new deception technologies and turn them into capabilities; (1) Without a clear understanding of the class of deceptions of interest, it will not be clear what capabilities would be desirable; and (2) Without a clear understanding of the objectives of the specific deception, it will not be clear how those capabilities should be used. If human deception is the objective, we can begin the planning process with a model of human cognition and its susceptibility to deception.

The skilled deception planner will start by considering the current and desired states of mind of the deception target in an attempt to create a scenario that will either change or retain the target's state of mind by using capabilities at hand. State of mind is generally only available when (1) we can read secret communications, (2) we have insider access, or (3) we are able to derive state of mind from observable outward behavior. Understanding the limits of controllable and uncontrollable target observables and the limits of intelligence required to assure that the target is getting and properly acting (or not acting) on the information provided to them is a very hard problem.

Deception Levels In the model depicted above and characterized by the diagram in Figure 2.6, three levels can be differentiated for clearer understanding and grouping of available techniques. They are characterized in Table 2.6 by mechanism, predictability, and analyzability:

Deception Guidelines This structuring leads to general guidelines for effective human deception, which are summarized in Table 2.7. In essence, they indicate the situations in which different levels of deception should be used and rules of thumb for their use.

Just as Sun Tzu created guidelines for deception, there are many modern pieces of advice that probably work pretty well in many situations. And like Sun Tzu, these are based on experience in the form of anecdotal data. As someone once said: *The plural of anecdote is statistics.*

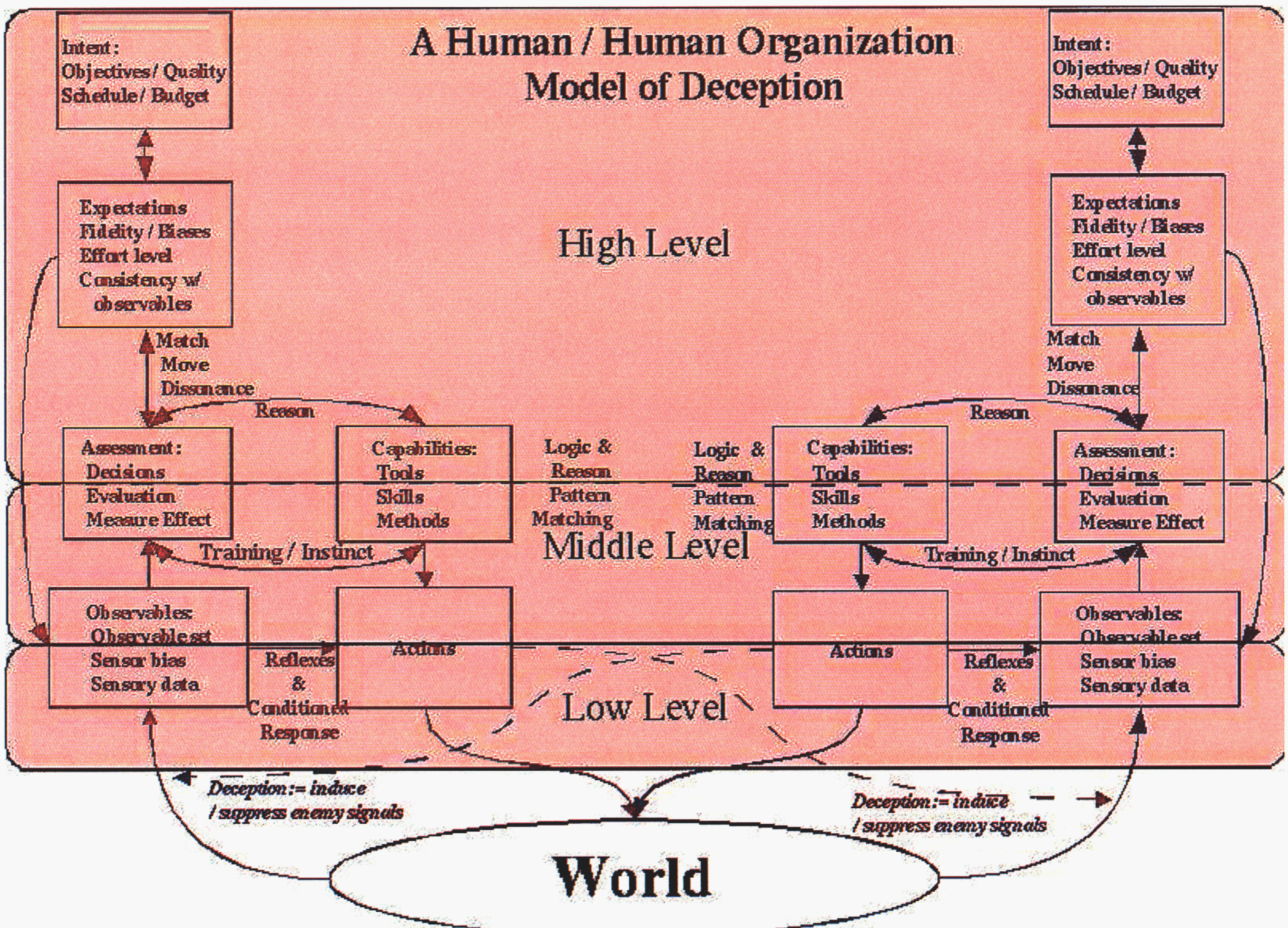


Figure 2.6: Human Model of Deception

Level	Mechanism	Predictability	Analysis	Summary
Low-level	Operate at the lower portions of the areas labeled observables and actions. They are designed to cause the target of the deception to be physically unable to observe signals or to cause the target to selectively observe signals.	Highly predictable based on human physiology and known reflexes.	can be analyzed and very clearly characterized through experiments that yield numerical results in terms of parameters such as detection thresholds, response times, recovery times, edge detection thresholds, and so forth.	Except in cases where the target has sustained physiological damage, these deceptions operate very reliably and predictably. The time frames for these deceptions tend to be in the range of milliseconds to seconds and they can be repeated reliably for ongoing effect.
Mid-Level	Operate in the upper part of the areas labeled Observables and Actions and in the lower part of the areas marked Assessment and Capabilities. Designed to either: (1) cause the target to invoke trained or pattern matching based responses and avoid deep thought that might induce unfavorable (to us) actions; or (2) induce the target to use high level cognitive functions, thus avoiding faster pattern matching responses.	Usually predictable but are affected by a number of factors that are rather complex, including but not limited to socialization processes and characteristics of the society in which the person was brought up and lives.	Analysis is based on a substantial body of literature. Experiments required for acquiring this knowledge are complex and of limited reliability. There are a relatively small number of highly predictable behaviors. These relatively small number of behaviors are common and are invoked under predictable circumstances.	Many can be induced with reasonable certainty through known mechanisms and will produce predictable results if applied with proper cautions, skills, and feedback. Some require social background information on the subject for high surety of results. The time frame for these deceptions tends to be seconds to hours with lasting residual effects that can last for days to weeks.
High-level	Operate from the upper half of the areas labeled Assessment and Capabilities to the top of the chart. They are designed to cause the subject to make a series of reasoned decisions by creating sequences of circumstances that move the individual to a desired mental state.	Reasonably controlled if adequate feedback is provided, but they are far less certain to work than lower level deceptions. The creation and alteration of expectations has been studied in detail and it is clearly a high skills activity where greater skill tends to prevail.	Requires a high level of feedback when used against a skilled adversary and less feedback under mismatch conditions. There is a substantial body of supporting literature in this area but it is not adequate to lead to purely analytical methods for judging deceptions.	A high skills game. A skilled and properly equipped team has a reasonable chance of carrying out such deceptions if adequate resources are applied and adequate feedback is available. Tend to operate over a time frame of hours to years and sometimes have unlimited residual effect.

Table 2.6: Deception Levels

Low-Level	- Higher certainty can be achieved at lower levels of perception. - Deception should be carried out at as low a level as feasible. - If items are to be hidden and can be made invisible to the target's sensors, this is preferred. - If a perfect simulation of a desired false situation can be created for the enemy sensors, this is preferred. - Do not invoke unnecessary mid-level responses and pattern matching - Try to avoid patterns that will create dissonance or uncertainty that would lead to deeper inspection.
Mid-Level	- If a low-level deception will not work, a mid-level deception must be used. - Time pressure and high stress combine to keep targets at mid-level cognitive activities. - Activities within normal situational expectations tend to be handled by mid-level decision processes. - Training tends to generate mid-level decision processes. - Mid-level deceptions require feedback for increased assurance. - Remain within the envelope of high-level expectations to avoid high level analysis. - Exceed the envelope of high-level expectations to trigger high level analysis.
High-Level	- If the target cannot be forced to make a mid-level decision in your favor, a high-level deception must be used. - It is easiest to reinforce existing predispositions. - To alter predisposition, high-level deception is required. - Movement from predisposition to new disposition should be made at a pace that does not create dissonance. - If target confusion is desired, information should be changed at a pace that creates dissonance. - In high-level deceptions, target expectations must be considered at all times. - High-level deceptions require the most feedback to measure effect and adapt to changing situations.

Table 2.7: Deception Guidelines

```
GIVEN: Deception A (low risk) and Deception B (high risk).
IF [A Succeeds] OR [B Succeeds] IMPLIES [Mission Accomplished, Good Quality/Sched/Cost]
AND [A Succeeds] AND [B Succeeds] IMPLIES
                                       [Mission Accomplished, Best Quality/Sched/Cost]
AND [A Discovered] OR [B Discovered ] IMPLIES [A (higher risk) AND B (higher risk)]
THEN DO B [comment: Do high-risk B first to insure minimal loss in case of detection]
    IF [B Succeeds] DO A (Late) [comment: Do low-risk A second to improve outcome]
        ELSE DO Out #1 [comment: Do higher-risk A because you're desperate.]
        OR ELSE DO Out #n [comment: Do something else instead.]

IF [A Succeeds] OR [B Succeeds] IMPLIES [Mission Accomplished, Good Quality/Sched/Cost]
AND [A Detected] OR [B Detected] IMPLIES [Mission Fails]
AND [A Discovered Early] OR [B Discovered Early] IMPLIES [Mission Fails somewhat]
AND [A Discovered Late] OR [B Discovered Late] IMPLIES [Mission Fails severely]
THEN DO B [comment: Do high-risk B first to test and advance situation]
    IF [B Early Succeeds] DO A (Late)
                     [comment: Do low-risk A second for max chance of success]
        IF [A Late Succeeds (likely)] THEN MISSION SUCCEEDS.
        ELSE [A Late Fails (unlikely)] THEN MISSION FAILS/in real trouble.
    ELSE [B Early Fails] [Early Failure]
        DO Out #1 [comment: Do successful retreat as pre-planned.]
        OR DO Out #m [comment: Do another pre-planned contingency instead.]
```

Table 2.8: Deception Algorithm

2.8.6 Deception Algorithms

As more and more of these sorts of rules of thumb based on experience are combined with empirical data from experiments, it is within the realm of plausibility to create more explicit algorithms for decision planning and evaluation. Here is an example of the codification of one such algorithm. It deals with the issue of sequencing of deceptions with different associated risks identified above.

Let's assume you have two deceptions, A (low risk) and B (high risk). Then, if the situation is such that the success of either means the mission is accomplished, the success of both simply raises the quality of the success (e.g. it costs less), and the discovery of either by the target will increase the risk that the other will also fail, then you should do A first to assure success. If A succeeds you then do B to improve the already successful result. If A fails, you either do something else or do B out of desperation. On the other hand, if the situation is such that the success of both A and B are required to accomplish the mission and if the discovery of either by the target early in execution will result in substantially less harm than discovery later in execution, then you should do B first so that losses are reduced if, as is more likely, B is detected. If B succeeds, you then do A. This is codified in Table 2.8 into a form more amenable to computer analysis and automation:

We clearly have a long way to go in codifying all of the aspects of deception and deception sequencing in such a form, but just as clearly, there is a path to the development of rules and rule-based analysis and generation methods for building deceptions that have effect and reduce or minimize risk, or perhaps optimize against a wide range of parameters in many situations. The next reasonable step down this line would be the creation of a set of analytical rules that could be codified and experimental support for establishing the metrics associated with these rules. A game theoretical approach might be one of the ways to go about analyzing these types of systems.

2.9 Summary, Conclusions, and Further Work

This paper has summarized a great deal of information on the history of deception in general and the historical, current, and emerging use of deception for information protection in specific. While there is a great deal to know about how deception has been used in the past, it seems quite clear that there will be far more to know about deception in the future. The information protection field has an increasingly pressing need for innovations that change the balance between attack and defense. It is clear from what we already know that deception techniques have the demonstrated ability to increase attacker workload and reduce attacker effectiveness while decreasing defender effort required for detection and providing substantial increases in defender understanding of attacker capabilities and intent.

Modern defensive computer deceptions are in their infancy, but they are moderately effective, even in this simplistic state. The necessary breakthrough that will turn these basic deception techniques and technologies into viable long-term defenses is the linkage of social sciences research with technical development. In specifics, we need to measure the effects and known characteristics of deceptions on the systems comprising of people and their information technology to create, understand, and exploit the psychological and physiological bases for the effectiveness of deceptions. The empirical basis for effective deception in other arenas is simply not available in the information protection arena today, and in order to attain it, there is a crying need for extensive experimentation in this arena.

To a large extent this work has been facilitated by the extensive literature on human and animal deception that has been generated over a long period of time. In recent years, the experimental evidence has accumulated to the point where there is a certain degree of general agreement in the part of the scientific community that studies deception about many of the underlying mechanisms, the character of deception, the issues in deception detection, and the facets that require further research. These same results and experimental techniques need to be applied to deception for information protection if we are to become designers of effective and reliable deceptions.

The most critical work that must be done in order to make progress is the systematic study of the effectiveness of deception techniques against combined systems with people and computers. This goes hand in hand with experiments on how to counter deceptions and the theoretical and practical limits of deceptions and deception technologies. In addition, codification of prior rules of engagement, the creation of simulation systems and expert systems for analysis of deceptions sequences, and a wide range of related work would clearly be beneficial as a means to apply the results of experiments once empirical results are available.

Chapter 3

Red Teaming Experiments with Deception Technologies

By Fred Cohen, Irwin Marin, Jeanne Sappington, Corbin Stewart, and Eric Thomas [1]
Draft of November 12, 2001

- Fred Cohen: Sandia National Laboratories
- Irwin Marin: The Emblematics Corporation
- Jeanne Sappington: The Emblematics Corporation
- Corbin Stewart: Sandia National Laboratories (CCD)
- Eric Thomas: Sandia National Laboratories (CCD)

3.1 Abstract

This paper overviews a series of 30 experimental runs designed to measure the effects of deception defenses on attacks against computer systems and networks.

3.2 Background, Introduction, and Overview

As part of an overall effort to understand the implications of technical deceptions in information protection, an effort was undertaken to perform experimental assessment of the use of specific deceptive methods against human attackers. This effort represents only a beginning down the path of understanding the role of deception in information protection, as outlined in Chapter 2. The specific set of technologies under study in this investigation were technologies similar to those described in earlier papers [Coh99a].

Because of the high cost in time and material of such a study, many goals were tied to this effort. They included: (1) improving the understanding of the participants in how systems are attacked and how they can be defended, (2) understanding how much an attacker can be told about a deceptive defense before they are able to defeat it, (3) understanding how deception impacts attacker workload, (4) understanding the group dynamics underlying attack groups and how it relates to success and failure, (5) understanding what sorts of ideas, strategies, and tactics arise in such groups when they

[1] This chapter is published online at `http://all.net/journal/deception/experiments/experiments.html`.

are not trained in any particular methodology of attack, and (6) understanding the impacts of initial access on the utility o deceptive defenses.

In total, 5 experimental runs of duration 4 hours each were run on each of 6 exercises. This represents 30 runs, including deception "on" and deception "off" control groups (6 each) and random "on" "off" mixes (18). Each run was preceded by a standard briefing and a run-specific briefing and followed by filling out of standard assessment forms, both individually by all team members and as a group. The exercises were of increasing intensity and difficulty so as to keep the participants challenged. Feedback was provided in the form of the exercise-specific briefing and was designed to first calibrate then systematically inform the attackers about more and more of the deceptive nature and type of the defense through the provision of 'intelligence' information being gathered by an insider. Eventually 'insider' access was granted to the attackers for measuring how they were able to perform with detailed knowledge and insider access to the nature of the deceptions. All experiments were repeated in very nearly identical circumstances with different groups of increasing suspected skill level and can be repeated again in separate runs for other groups. A few of these experiments were repeated with higher quality attack groups with.

3.3 The Laboratory Environment

The laboratory environment used for these red teaming experiments consisted of two rooms.

- The first room is used by the attackers for their attacks. It consists of a set of attack computers and research computers. (1) The research computers are designed to provide the attackers with access to Internet and previously prepared capabilities and techniques as well as to provide access to additional computing capabilities, databases, and other individuals they may wish to seek help from. (2) The attack computers are configured in known configurations and are designed to facilitate attacks of the sorts known to the attackers. The attackers are permitted to, and often do, bring their own system capabilities to the exercise. Systems in this room are instrumented to allow attack methods to be reviewed later and the room has a videotape machine for taping sessions. It also has a computer used by the observer to take notes, is separated from the rest of the laboratory, and allows external access for bathrooms and other needs.

- The second room houses the systems under attack. It is physically separated from the attack room and is locked to prevent attackers from accessing it. It includes a set of systems and wiring capabilities to allow any network containing less than a few dozen computers to be rapidly configured and reconfigured to facilitate experiments.

The cost to supply such a laboratory is on the order of $40,000, most of which is in the cost of equipment. It took on the order of 50 person days to create the environment, In the case of these experiments, the laboratory itself is reasonably physically secure and has additional protections in this form of digital diodes to assure that information from experiments does not leak to the rest of the world. This is intended to assure that attacks do not spill over into the general Internet. A reasonable estimate of the costs of repeating these experiments would have to include the cost of labor (6 people for 5 hours for each run plus analytical time and experimental design and configuration time, and other support) comes to approximately $1000 per experiment plus $1,000 per run, or about $54,000 for this set of experiments. Facility space, electrical power, and other overhead bring the total cost of such an experiment to something on the order of $150,000.

3.4 Repeatability in Experiments

In order to assure essentially repeatable experiments, there are a set of file servers used to store complete disk images of experimental configurations. Using the Samba protocol and a bootable CD-

ROM, we are able to make forensically sound images of systems to be attacked and systems launching attacks before and after experiments. The pre-experiment images are reloaded into experimental systems prior to each experiment so that all systems involved in the experiment are, in essence, identical. The one exception is that experiments are run on different days, and sometimes at different times of day to accommodate schedules.

The ability to create a very nearly identical experimental environment is critical to such research and there is a considerable cost associated with this. For example, even at relatively high network speeds, it costs on the order of 12 minutes per system to make an image and another 12 minutes to restore that image. This means that reproducing an experiment requires something like an hour of preparation time as well as possible network reconfiguration.

All experiments are permanently archived so that they can be repeated at a later date and time by the same group or another group of test subjects. This allows effects like training, experiment order, and subject biases to be remediated and allows groups to repeat experiments after training, after being provided with additional information, and after intentional introduction of biases.

3.5 Effects Under Consideration

In the initial 45 experiments performed in this environment we were most interested in several primary factors:

- The difference in performance with and without deceptions in place is fundamental to our desired understanding. In order to observe this effect, open ended exercises are used. In these sorts of efforts, the problem is sufficiently complex for the time provided that it would be an exceptional team that could complete all facets of the challenge in the allotted time. The experiments have sets of goals that, in essence, require the achievement of some earlier objectives to achieve some later objectives. The objective of deceptions in this case is to reduce the effectiveness of attackers. The metric is then how far the attackers get how fast rather than their ability to complete all tasks. In this sense, the problems are like mazes without end and the characterization we use to describe them later is an attack graph. A fully successful attack would, presumably, have to follow one of a small number of attack graphs that lead to success. Other graphs lead to false success (when deception is in place) or to failures or delays. We can then measure success relative to finding one of the paths that leads down a successful attack graph.

- The difference in performance of attackers between situations when the deception is known to the attacker and when it is unknown to the attacker was also vital to our understanding because we were interested in the performance of deceptive defenses in the presence of insider threats, intelligence threats, and overrun threats. Thus we performed experiments with different levels of knowledge provided to the attackers so that we could measure the performance difference based on their knowledge of the situation.

- Based on some initial theoretical work we believe that there may be a correlation between success and the types of deceptions we are trying to induce. Specifically, we sought to differentiate deceptions that induce type 1 (omission), 2 (commission), and 3 (misdirection) errors and to understand he thresholds at which these types of errors occur, are detected or suspected by attackers. and can be induced with effect.

In this initial experiments, only these three factors were explored, however, we are also interested in aspects of the nature of deception, as described in Chapter 2, and the way in which they operate in the information defense arena. Specifically, we are interested in how limited resources lead to controlled focus of attention, how effective deceptions can be composed from concealments and simulations, how memory and cognitive structure force uncertainty, predictability, and novelty and how this can be exploited for deception, how time, timing, and sequence work in deceptions, how much control over observables are required, operational security requirements, effects of different attack methodologies and capabilities, the recursive nature of deceptions, how small changes can impact large systems, the complexity required for implementing deceptions to great effect, what level of knowledge of the target is required to be effective over what time frames, how deceptions can be modeled and outcomes predicted, and how counterdeception functions.

3.6 Additional Goals of Exercises

As part of these exercises, we also hoped to advance the knowledge and skills of the participants. The participants, in this case, were students ranging in age from 16 to 38, all in computer-related fields, all with excellent grade point averages, all US citizens, and all interested in information protection, and all participating in an intensive program of study and research in this area. Through this effort, we hoped to give them skills and knowledge that would be helpful in understanding how systems are attacked and how they may be more effectively protected. The students were also taught classes on information protection, received training in how to manage and operate systems, and participated in hands on research and systems administration projects over the period of this effort.

The same exercises were also run on more skilled attackers including teams of professionals that do testing of high assurance systems, professional red teaming groups, professionals in the field of information system intelligence, and professional offensive information warriors. These experiments are used to calibrate the results. This paper does not include these results in its findings because they were not statistically meaningful, however, they were consistent with our other results.

3.7 Summary of Collected Data

The collected data consists of evaluation forms filled out by all participants after each session, a group form filled out as a consensus in a facilitated group meeting after individual forms were completed, a summary of events and times as recorded by the observer, and detailed copies of the system configurations before and after each exercise[2]. Standard pre-briefings were provided for each group to assure to a reasonable extent that groups would keep results independent of each other and to provide reasonable limits on behavior while fulfilling administrative requirements of the facility (see Appendix A. Forms were designed so as to solicit specific information related to research interests (see Appendix B). Specifically, questions were directed toward determining whether deceptions were thought to have been identified and bypassed, understanding whether the participants were operating in level 1, 2, or 3 of the cognitive characterization used in the framework for deception in Chapter 2 which forms the basis for this work, detecting issues in group behavior that relate to success and failure of deceptions (e.g., the effect of the group on preventing exploration of lines and the effect of the group on inducing lines), and information on the strategies employed and tools use and effectiveness, which are directed at improving performance of other groups in similar tasks.

After each set of experiments, full details were provided to all participants. Thus the sequencing of experiments went from (1) no revelation of deception issues to (2) provision of details about the presence of deceptions and the deception technologies in use and finally (3) to full details of the deceptions including all configuration details. This enabled us to measure across the dimension of knowledge of the deception. Control groups were used with deception always off and deception always on so that cross-experiment differences in time to achieve goals could be measured. These groups were maintained within each sub experiment (3 weeks duration) but groups were reshuffled after each three week period to try to find group mixes that tended to improve performance on red teaming efforts and to help students learn how to work well in project groups and learn more from each others talents and skills.

[2]Detailed research data is not available for reasons of participant confidentiality.

There were also faults detected in experiments. While we do not believe that any of these faults invalidate the overall results, additional experiments and improved experimental conditions would be helpful in mitigating such faults in the future. Specifically, fault fell into the following categories; (1) limits of the facilities and situation, (2) limits of the experimenters and time frames, (3) limits of the technology employed, (4) experimenter and participant error.

- **limits of the facilities and situation** The facilities were being upgraded and altered under us while these experiments were being performed and the facility was never intended for this sort of experiment. Interruptions were kept to a minimum, but they did occur, a network outage interrupted the location of Internet data on one occasion, the technology used to facilitate the work was less than ideal, and there were days without air conditioning when it was over 80 degrees Fahrenheit in the attacker's facility. We did all we could to keep things equitable, but clearly these sorts of conditions have some impacts on performance.

- **limits of the experimenters and time frames** The experimenters involved were not professionals in this realm and thus were not perhaps as good as their jobs as some others might not have been. In addition, it was necessary for the observer to have knowledge of the real situation and to be in the same room as the subjects. Thus there was the potential for bias and, on some occasions, there was laughing by observers and interaction between subjects and observers. The time frames for setting up and running these experiments were also very tight, so experiments did not always function perfectly and imperfections observed by the observer were repaired while the experiment was ongoing. while efforts were made to avoid any direct information from this activity, on several occasions subjects suspected that the observer had altered the experiment.

- **limits of the technology employed** The specific deception technologies employed were thrown together on very little notice, as was the environment for the deceptions. This was because of the short window of opportunity to collect data while there were enough subjects available. This caused a variety of complexities, but for the most part, the same conditions were present for each group so that these issues tended to even themselves out.

- **experimenter and subject error** These included cases where experimenters and participants made various mistakes. In particular: (1) We had two cases where a subject indicated that they had reversed the meanings of the numerical values in the evaluation forms during the out briefing when all participants were asked to come up with numerical values together. We corrected the values in these subject's forms immediately thereafter by inverting the values (5 became 1, 1 became 5, and so forth). (2) In one experiment an error in system configuration prohibited progress for more than an hour. This was mitigated during the experiment and the time difference between the time the same activity that showed the error and the time when it was compensated for was subtracted from subsequent times in the results. (3) In a few cases the familiarity of the subjects with the observers, the presence of additional observers, or the presence of a camera in the room caused limited interference with the experiments, however, we do not believe that these had any effects on the progress relative to the attack graph from a standpoint of differences between the presence and absence of deceptions. Specific cases are noted below where appropriate.

Finally, as in many such experiments, the subjects were predominantly academically skilled college students studying computer security at a national laboratory. While these results look promising, such students almost certainly represent only a small segment of the space of real attackers, and are far less skilled than many real attackers. Select experiments were also performed with other groups and details are provided for those cases below.

It would clearly be desirable to repeat these experiments under more realistic conditions, however, we do not believe that these conditions had any serious impact on outcomes and we believe that money spend on such efforts would be better spent doing other experiments which provide additional results while covering the issues in this set of experiments as a side effect of those efforts to detect any refutations should they arise, or to provide confirmations of these results.

3.8 The Structure of Attack Graphs

In each experiment, there were known successful attack graphs and actual attack graphs followed by participants. In Figure 3.1 through Figure 3.4, we summarize the successful attack graphs for each run, so that they can be compared to actual attack graphs, and alternative attack graphs yielding type 1, 2, and 3 errors, as observed in experiments. Unlimited numbers of additional attack graphs are likely feasible for successful attack, seemingly successful attack (deceptions effective), and failed attacks.

While these high level representations of attack graphs are not strictly accurate to the details of attack sequences, they are helpful in understanding the nature of the situation. Metrics could reasonably be related to each link in these graphs with the resulting weighted graph providing measures for the difficulty of attack given the deception situation. Creating these weights requires two things. (1) There are strictly mathematical issues, such as the number of paths in some direction and their distribution, that might lead to purely mathematical values for some metrics. For these direct solutions can be applied. (2) The rest of the situations depend on the relative skill of the attacker in detecting the victim and differentiating it from the deceptions. This detection and differentiation problem comes down to peoples' ability to devise automation and use their own analytical capabilities. This sort of data can only be found through empirical measurement, or in other words, experiments.

3.9 Actual Graphs Followed

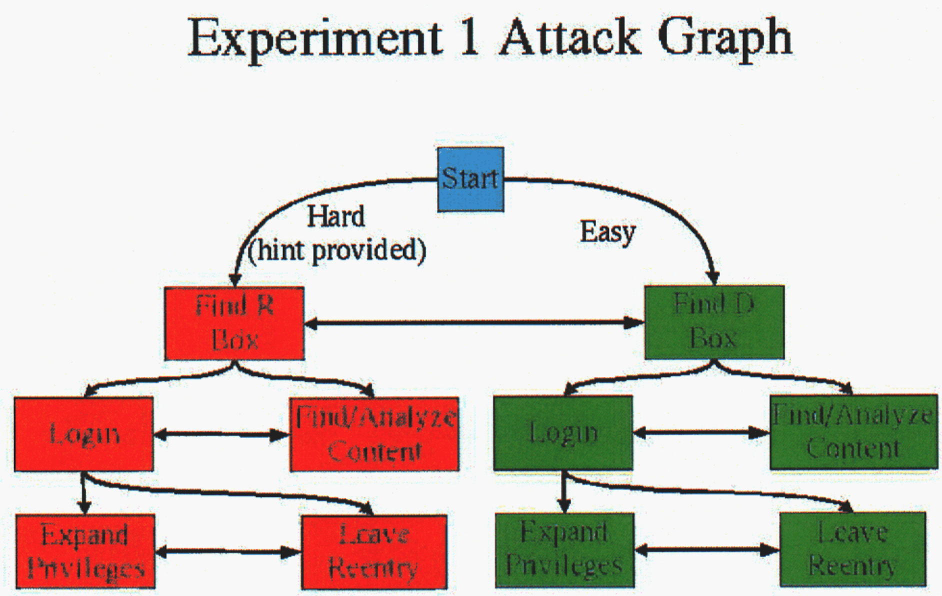
Each group in each experiment followed an actual attack graph over time. These attack graphs are summarized here along with some interpretation in Table 3.1. We use the term "Hop" interchangeably with "Experiment" and indicate the first time the attacker got to any given step (in the case of some deception systems, steps may have to be retried many times).

The plot in Figure 3.5 plot summarizes this data in a different format. In this summary, each run is represented by a line. Lines in red indicate attack sequences with deception enabled while lines in blue show attack sequences with deception disabled. The 'X' axis represents time, while the 'Y' axis is positive for 'Real' locations in the attack graph and negative for 'Deception' locations in the attack graph.

If anything is clear from this plot it is that attackers do better without deception. This is no surprise. However, there are a lot of other interesting characteristics in these results that we will now discuss. The table in Appendix C summarizes detailed information on factors identified for measurement in the experiment and called out in the provided forms.

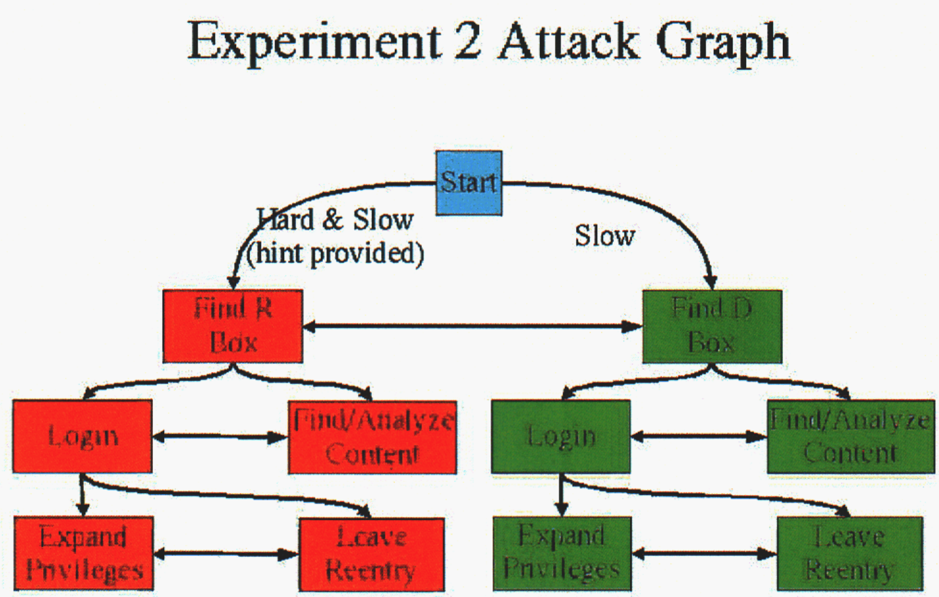
3.10 Analysis

The first and perhaps most important thing to notice in the summary of results is that when deception is enabled, attackers never get as far toward the truth as they do when deception is disabled. In other words, deception works. Furthermore, it works very well. When deception is turned on, attackers almost uniformly go down the deception parts of the attack graphs rather than down the real parts of the attack graph. In cases other than blatant dazzlement, they are convinced that they are going down real paths for a substantial time. In some cases, attackers were so convinced that they had won when they were actually deceived, that they declared victory and walked away early. In some dazzlement cases, people got so frustrated that they gave up early.



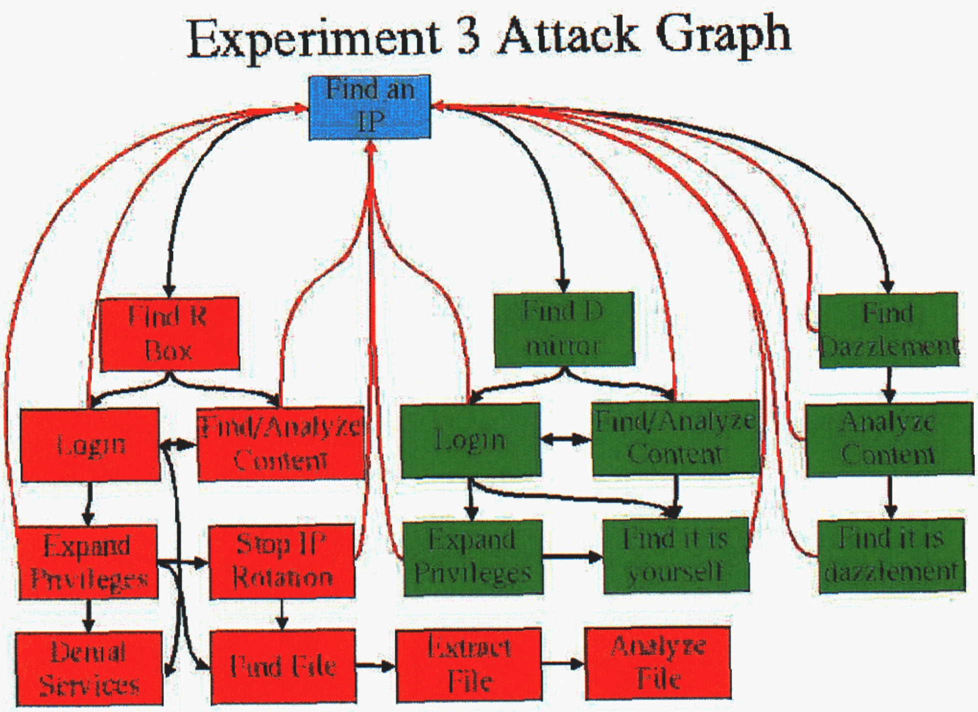
- **1R or 1D** find box (easy) - D directs target to wrong victim
- **2R or 2D** log in — find content (Wrong path looks good)
- find content — analyze content (Wrong path looks good)
- analyze content — login (Wrong path looks good)
- **3R or 3D** leave reentry — expand privileges (Wrong path looks good)
- expand privileges — leave reentry
- **4R or 4D** target believes they win when they lose and deceiver observes and learns about target

Figure 3.1: Experiment 1 Attack Graph



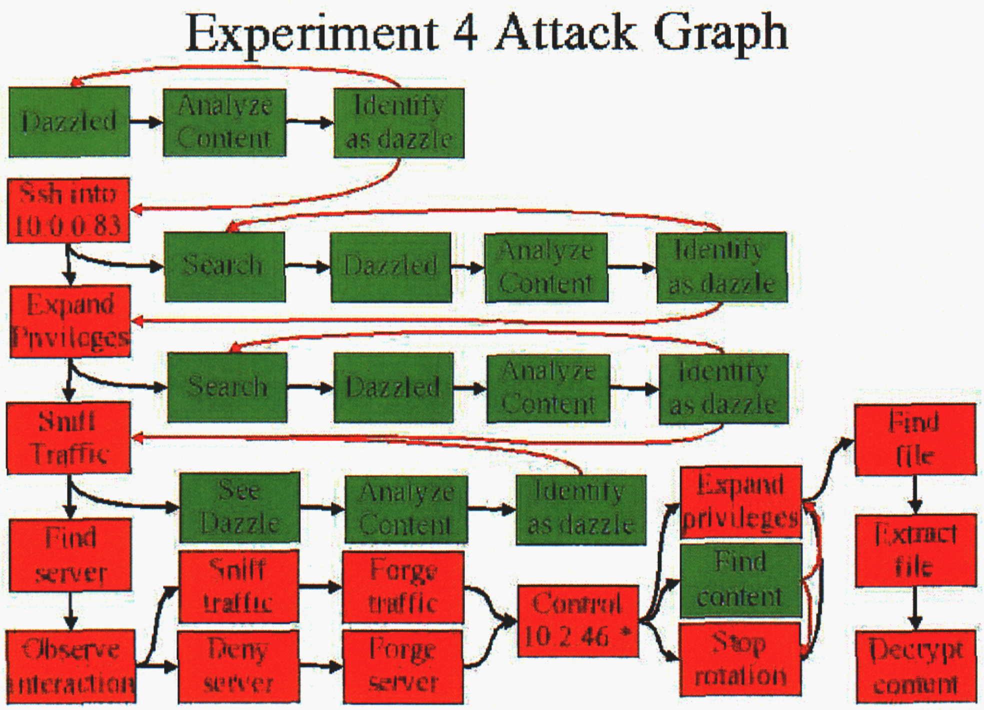
- **1R or 1D** find box (hard) - D directs target to wrong victim, search is very slow, time pressure induces alternative search strategies, some search strategies reveal deception - but are not noticed
- **2R or 2D** log in — find content (Wrong path looks good)
- find content — analyze content (Wrong path looks good)
- analyze content — login (Wrong path looks good)
- **3R or 3D** leave reentry — expand privileges (Wrong path looks good)
- expand privileges — leave reentry
- **4R or 4D** target believes they win when they lose and deceiver observes and learns about target

Figure 3.2: Experiment 2 Attack Graph



- **1R or 1D** loop: find box - Deception makes differentiating box harder and increases find (real) box time dramatically
- **2R or 2D** log in — find/analyze content (Wrong path consumes time)
- Addresses change before success ⇒ goto loop
- Trigger detector ⇒ goto loop w/shorter times
- **3R or 3D** time low ⇒ deny services - but deny to what? - and tell how?
- **4R or 4D** leave reentry — expand privileges
- leave bug ⇒ easier to find
- **5R** stop movement ⇒ easier to find — plant Trojan ⇒ easier to find
- **6R** find file
- **7R** extract file and analyze file

Figure 3.3: Experiment 3 Attack Graph



- **1D** Search for or try to analyze 10.0.0.83 and ignore intelligence provided
- **1R** Enter 10.0.0.83 via ssh
- **2D** Search for other systems in 10.0.*.* and try to exploit them
- **2R** Expand privileges using routine provided
- **3D** Search network for targets to attack
- **3R** Sniff traffic
- **4D** See dazzlement, analyze, identify as dazzlement
- **4R** Find real client and server and observe traffic
- **5R** Understand interaction and determine a viable attack
- **6R** Gain control of the victim
- **7D** Look for content (unfindable in this state)
- **7R** Expand privileges
- **8R** Find file
- **9R** Extract file and analyze content

Figure 3.4: Experiment 4 Attack Graph

Group	Hop	D	Step/Time	Step/Time	Step/Time	Step/Time	Step/Time	Step/Time	Step/Time
Mon	1	No	1R 2:00	2R 2:08	3R 2:45				
Tue	1	No	1R 0:22	2R 0:24	3R 1:11	4R 3:27			
Wed	1	No	1R 1:58	2R 1:58					
Thu	1	Yes	1D 0:17	2D 0:20	3D 0:22	4D 2:26			
Fri	1	Yes	1D 0:31	2D 0:31	3D 3:08	4D 3:23			
Mon	2	Yes	1D 3:37						
Tue	2	No	1R 3:33						
Wed	2	No	1R 1:37	2R 1:42					
Thu	2	Yes	1D 1:48 *	2D 2:06					
Fri	2	No	1R 0:40	2R 0:49					
Mon	3	No	1R 0:41	2R 1:25					
Tue	3	No	1R 1:15	2R 2:58					
Wed	3	Yes	1D/R 0:52						
Thu	3	Yes	1D/R 0:17						
Fri	3	No	1R 0:29	2R 0:51					
Mon @	4-1	Yes	1D 0:38	1R 2:07	2R 2:16	2D 3:01	3D 3:20		
Tue	4-1	No	1D 0:30 *+1*	1R 0:45 *+2*	2D 0:50	2R 1:40	3R 1:45	4R 1:50	
Wed	4-1	No	1R 0:21	1D 0:30	2R 0:42	3R 1:05	4R 1:30	5R 2:45 *+3*	
ThuA	4-1	Yes	1R 1:34	2R 1:45					
Thu	4-1	Yes	1D 0:55	1R 1:35	2R 1:50	3D 2:23	1D 2:23 *+4*	3D 2:55	
Fri	4-1	Yes	1R 0:37	2R 0:54	3D 1:43	1D 2:31	4D 2:43	3R 3:37	
Mon+	4-2	Yes	1D 0:51	1R 1:32	2R 1:41	3R 1:45	4D 1:45		
Tue+	4-2	No	1D 0:34	1R 1:22	2R 1:33	3R 2:10	4R 2:10		
Wed+	4-2	No	1R 1:45	2R 2:18	3R 2:30	4D 3:12 *+5*			
Thu+	4-2	Yes	1R 0:47	2R 0:58	3D 1:12	3R 3:15			
Fri+	4-2	Yes	-	-	-	-	-	-	-
Mon+	4-3	Yes	1R 0:20	3R 0:59	2D 1:45	2R 1:59	3R 2:06	3D 2:22	4R 3:01
Tue+	4-3	No	1R 0:27	2R 0:28	3R 1:10	4R 1:24			
Wed+	4-3	No	1R 0:18	2R 0:19	3R 0:23	4R 1:32	5R 3:10		
Thu+	4-3	Yes	-	-	-	-	-	-	-
Fri+	4-3	Yes	-	-	-	-	-	-	-
SR-1 *+6*	3.1	Yes	1D/1R	2D/2R	3D/3R				

***** They achieved 1R at 2:06 but never realized it because they were occupied with following the line of 2D.

@ Groups re-aligned after Run 3. Teams briefed on the deceptions and technologies in use.

+1 Even with deception turned off, teams try various lines that are not fruitful. They did not observe a deception, which accounts for rapidly moving to 1R. On the previous day, the deception caused about 1.5 hours of delay.

+2 Due to an experimental fault 1:45 was wasted between 0:30 and 0:45, so times have been adjusted backwards to reflect progress toward the goal.

+ Experiment 4 was run three times on the same groups to give them more opportunity to spend more time on the same problem, including the development of improved tools.

+3 They see the interaction but do not yet realize what it really is.

A Additional exercise in the morning (AM) for 4 hours, involving the team that designed the experiments (but not the person who built the specifics of this run).

+4 They lose confidence in the real line because of dazzlements (3D) and return to 1D believing the original dazzlement over the real system they were in.

+5 they do not differentiate their own scans and deceive themselves temporarily.

+6 This was an 'extra run' of a slightly enhanced experiment 3. Details are provided below under 'special runs'.

- indicates a team that decided not to participate.

Table 3.1: Actual Attack Graphs

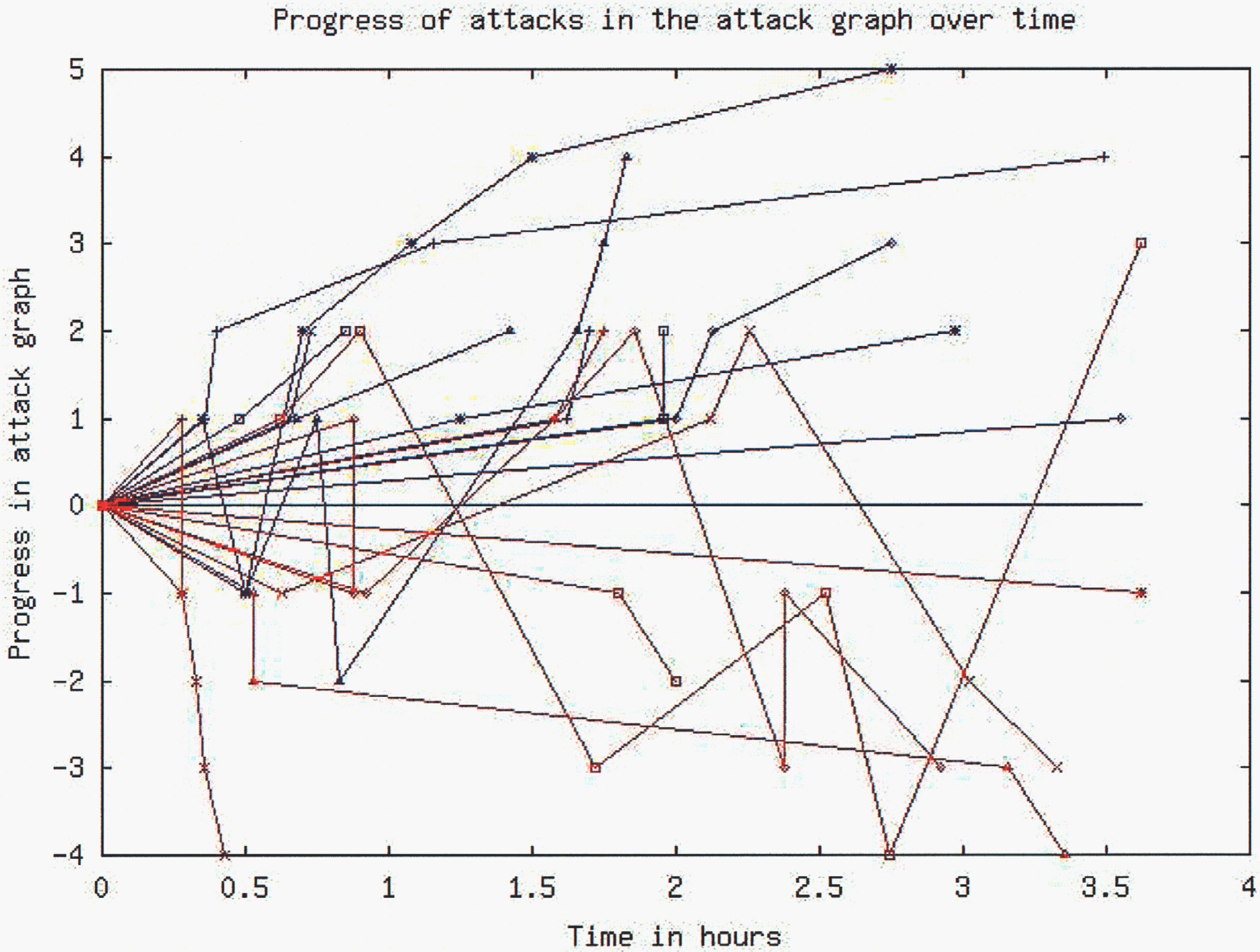


Figure 3.5: Progress of Attacks over Time

These results verify the previous anecdotal data from the HoneyNet project [Hon] and Deception ToolKit [Coh98a].

3.10.1 The First Four Weeks of Experiments

In the following section, we examine the plots for each of the first four weeks of experiments, one week at a time. In the first three weeks, teams put earlier in the week were thought to be less able based on their known skill sets, no training was done for any teams, and a control group for each of non-deception (Tuesdays) and all-deception (Thursdays) were provided. This gave an advantage to the control group with deception enabled because they were considered more highly skilled than the control group with deception disabled. Despite this advantage, deception was quite successful.

In the first week (Figure 3.6), we induced type 3 errors - errors of misdirection. This was done by introducing two identical systems, one obvious and available, and the other less obvious and harder to find. When deception was turned on, the obvious system was the deception while the in-obvious system was to real object of the attackers' efforts. When deception was turned off, the only system to find was easily found and exploited. When deception was turned on, the obvious system was easily found and exploited. Since it appeared to the attackers that they had won, they gave up the effort and declared victory. The plot makes this very clear in that all of the non-deception attacks (in blue) succeeded, while all of the deception attacks (in red) failed but appeared to succeed.

In the second week (Figure 3.7), we induced type 2 errors - that is - errors of commission. This was done by providing a very large search space that was very thinly populated with real targets and, when deception was enabled, equally populated with easier to find deception systems. These experiments were designed to be open-ended in that it would be very hard for any team to get to the end in the time allotted. While no team got very far in these experiments, it is clear that the teams that ran without deception enabled made progress in the right direction while teams working with deception enabled made nearly the same amount of progress, but in the wrong direction. In other words, the rate of progress through the attack graph was similar but the deception was extremely effective.

In the third week (Figure 3.8), we induced type 1 errors - that is - errors of omission. This was done by providing an enormous quantity of false data - to the point where detecting possible targets was very easy but differentiating real targets from false ones was far harder. As we can clearly see, the effect of this deception was that attackers found both the deceptions and the real systems at about the same time but were unable to differentiate them and unable to make further progress. The attackers not undergoing deception were able to identify the real victim systems and make progress against them.

Week four (Figure 3.9) represents the beginning of the second phase of the effort. For the first three weeks, the attackers under study were not told anything about deception or its use. While some of them guessed that deception was in use when they saw dazzlements, no information was provided from those performing the study. At the end of the third week, all groups were out briefed on the nature of the experiments and given full details about the nature of the research, the deceptions that were used, how they performed, and the technologies that were used. Starting in week 4, two changes were thus made. The first change was that the information provided about the situation included information about deceptions. The second change was that the next three weeks of attacks were performed against the same defenses. Thus teams could learn from previous efforts and attempt to avoid the mistakes of the previous week in the next week. Week four can thus be considered both in the context of the previous three weeks of experiments and in the context of the subsequent two weeks of experiments.

In week four, the system under attack was more complex than in previous weeks and for that reason we expect that total progress relative to the ultimate objective will not be as rapid. In fact the group that made the best progress resulted in a failure that could not be recovered from (the attackers caused an intermediary system required for success to fail). Interestingly, we see an increased variety of effects when the teams understand that deception is a possibility.

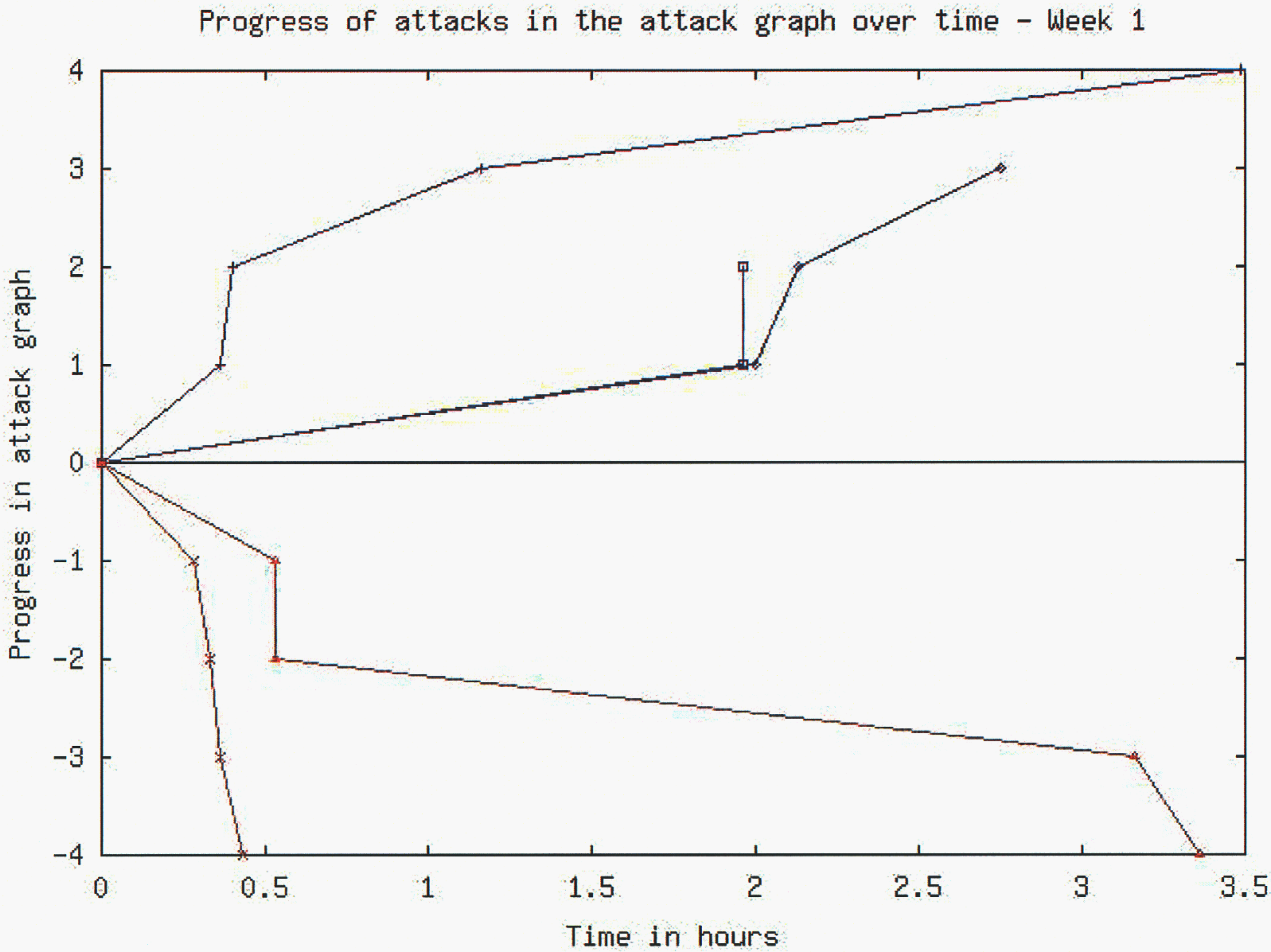


Figure 3.6: Week 1 Progress of Attacks

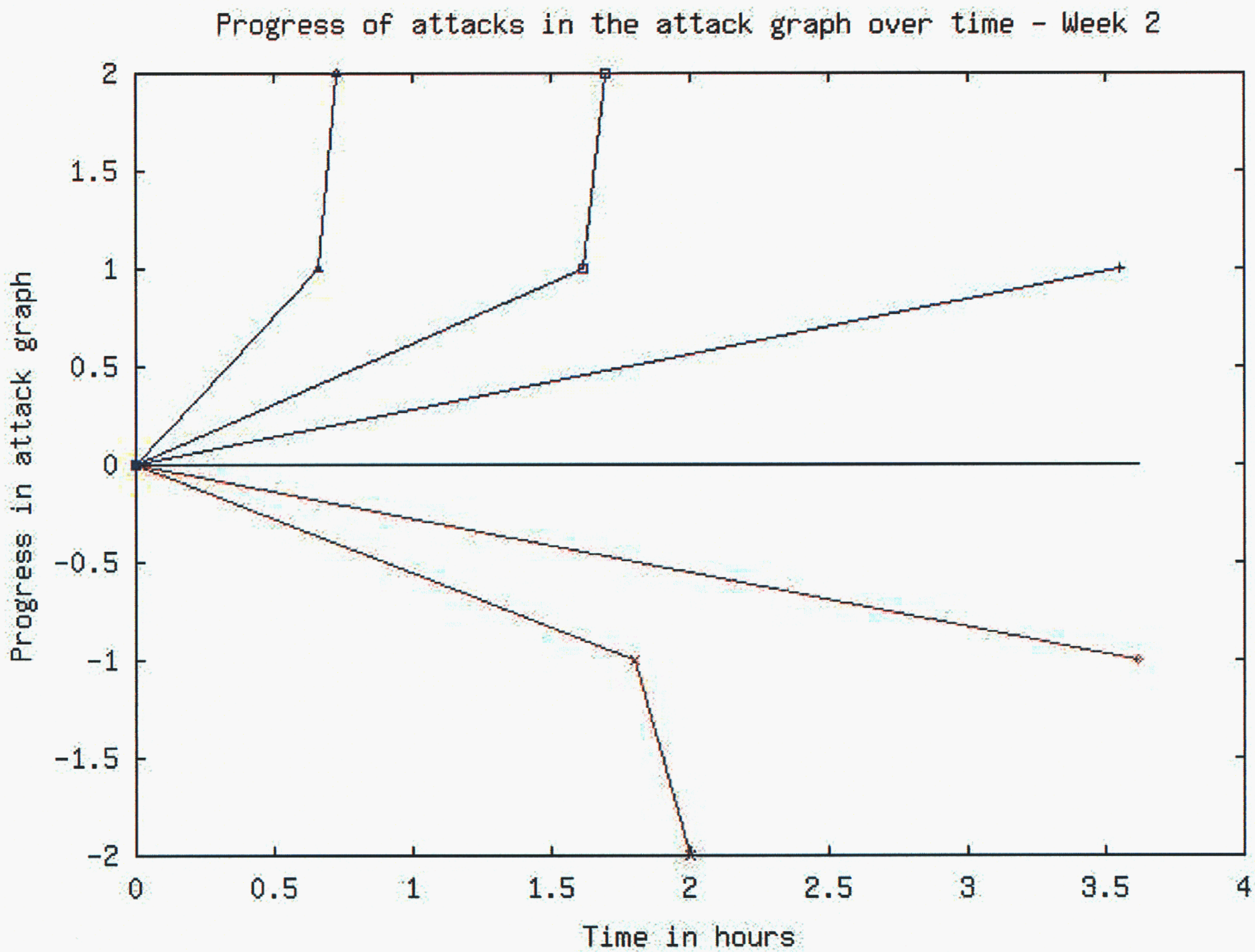


Figure 3.7: Week 2 Progress of Attacks

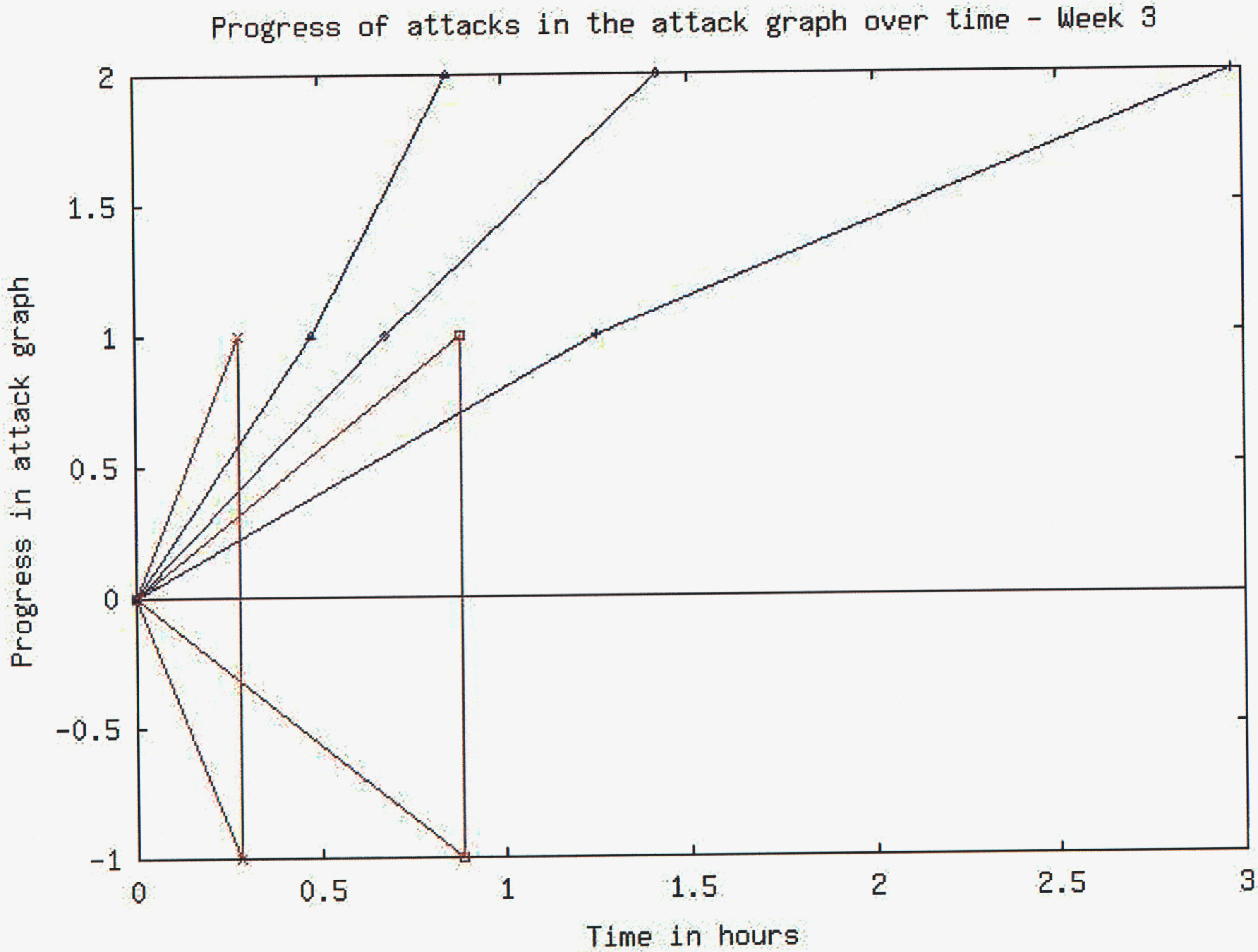


Figure 3.8: Week 3 Progress of Attacks

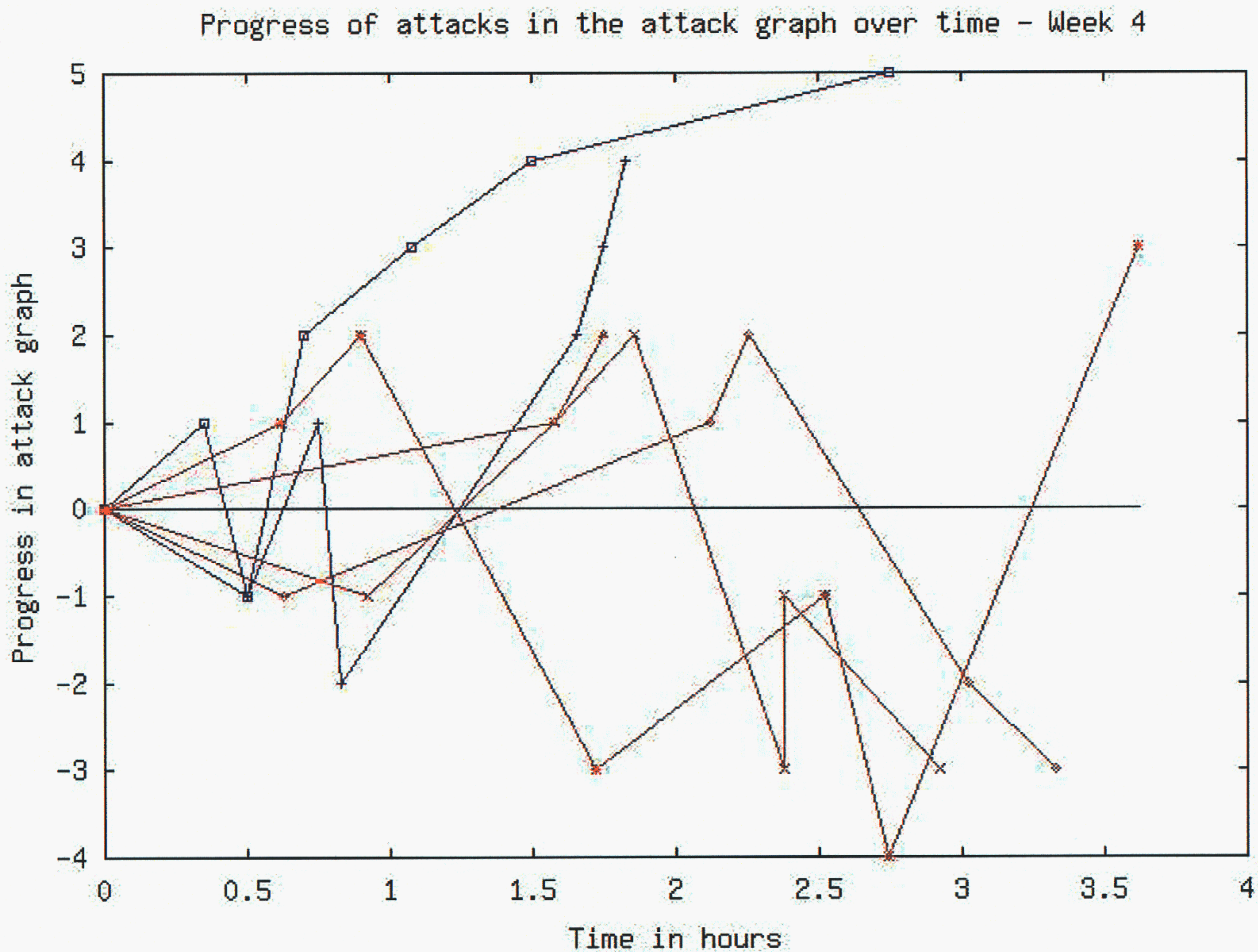


Figure 3.9: Week 4 Progress of Attacks

- One of the most startling effects is that teams suffer from self-deception. For example, the two teams that were not being deceived believed that they were being deceived at various times and acted on those self-deceptions. They performed additional experiments similar to those that someone being deceived would attempt and thus we noted these as deceptions in the plot. They recovered fairly rapidly in comparison to teams actually being deceived, but this indicates that the mere threat of deception offers some protective value.

- A sixth team participated in this week's activities as well. This team consisted of the people who designed the experiments and included some of the people who had watched previous teams in these same exercises and who had almost complete knowledge of the manner in which the experiment was being undertaken. They had been previously briefed on the attack graphs including the deception paths and were extremely cautious in their approach. They included a senior intelligence officer (recently retired), two highly skilled systems administrators, a naval researcher, and a highly skilled security consultant who used to run intelligence operations for a state law enforcement agency. This group did not encounter any deceptions, and they made slow but steady progress toward their goal. Because of time limitations on the facility they had one hour less than the other teams and got further in the time allotted than the other two teams exposed to deceptions. The left very little in the way of footprints of their attacks, and while it is likely that they would have encountered deceptions in their next step, their experience and knowledge of the detailed attack graphs clearly benefitted them. They did not, however, progress as far as the far less experienced teams that were not facing deceptions.

- Backtracking behavior was encountered among groups that were being deceived, and this resulted in revisiting parts of the attack graph that had previously been encountered and being (in one case) re-deceived or (in the other case) deceived by a deception that had previously been avoided. The fist case is seen in the team that achieved -1 at 1 hour, -3 at 2.4 hours, and -1 again at 2.4 hours. The second case is seen where another team encounters -3 at 1.7 hours and then encountered -2 for the first time at 2.5 hours.

- The movement back and forth between real progress and false progress, between reality and deception, and between deception closer to and further from the starting point indicate that measuring progress toward the goal is far more difficult for the targets of the deception to assess because of the lack of clear and consistent feedback available by direct observation. The problem of counterdeception is clearly in play here and the need for some high assurance feedback for the attackers seems clear if progress is going to be made against such deceptive defenses.

3.10.2 Confounding Factors in the First Four Weeks

In our Chapter 2 we identified a set of confounding factors associated with deception. Specifically, these are factors that affect movement between the three levels of cognition (low-level, mid-level, and high-level) identified in the previous cognitive model. The questionnaire that team members filled out after experiments and then filed out as a group combined with the observer's notes were intended to allow us to measure these factors. The data on confounding factors is analyzed to understand the relationship between these factors and performance.

The first summary, shown in Table 3.2, indicates that the difference between results for all confounding factors when deception was on and off was fairly insignificant. The largest difference was 0.49 and this with a standard deviation of more than 1.1. In all cases the difference (shown in the last column) is less than 50% of one standard deviation. From this we can conclude that there was no clear relationship between the presence or absence deception and any of the identified confounding factors in these experiments. Strangely, time pressure and distraction were lower when deception was on than when it was off. While these differences are not very statistically significant, they appear to be present.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time
Off	3.31	3.06	2.61	2.96	2.71	2.61	3.76	3.14
StdDev	1.08	1.34	1.22	0.87	0.89	1.15	1.15	1.1
On	3.15	2.83	2.63	2.73	2.73	2.6	3.68	2.65
StdDev	0.77	0.93	1.17	0.78	0.93	1.1	0.94	1.08
Diff	0.16	0.24	-0.01	0.23	-0.01	0.01	0.08	0.49

D	Uncert	Distract	Tired	Hard	Int	Joy	Surp
Off	3.16	2.73	2.92	3.94	3.57	3.2	3.41
StdDev	1.05	1.22	1.17	0.94	1.02	0.87	0.79
On	3.43	2.35	2.75	3.9	3.58	3.23	3.28
StdDev	0.98	0.89	0.81	0.93	1.15	1.12	0.85
Diff	-0.26	0.38	0.17	0.04	0	-0.02	0.13

Table 3.2: Relationship Between Deception and Confounding Factors

If we examine the same data week by week (Table 3.3), we find that, despite a relatively small total amount of data per category, there are only a few cases where the difference in mean between deception enabled and deception disabled exceeds one standard deviation.

- In week 1 and week 4, both original and new strategies reportedly worked better with deception off than with deception on, but this result did not hold across weeks 2 and 3 and the statistical significance is between the 'on' and 'off' standard deviations in both cases where it appears.
- The belief in success was higher in week 1 when deception was present than when deception was not present, but in week 4 deception caused reduced belief in success. The deception in week 1 was very effective in concealing the fact of deception, while in subsequent weeks deception was not only more obvious, but also believed and eventually known to be present. The importance of success fell off in week 4, but this will be discussed in more depth in the analysis of weeks 4-6 below.
- Time pressure was perceived as higher with deception on than off in week 1, but this did not hold for later weeks.
- Uncertainty was higher for weeks 1 and 3 with deception on, but not in weeks 2 and 4, and not to a very significant extent. Distraction was negatively correlated with deception in all four weeks, but not at a very significant level.
- Exhaustion was never an issue, but difficulty was believed to be lower in weeks 1 and 2 when deception was enabled, while it was higher in weeks 3 and 4 when deception was enabled. This may be related to the suspicion and eventual knowledge of the presence of deception that grew over time.
- Increased difficulty was somewhat correlated to increased interest and in week 3, interest was higher when deception was on, but generally interest was kept high throughout these four weeks of experiments.
- Enjoyment was negatively correlated to deception in all except the third week, where the increased interest and difficulty apparently drove the subjects to desire to meet the challenge.
- No significant difference in surprise correlated to deception was reported in any of the experiments.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time
Week 1 On	3	2.88	3.75	3.25	3.63	4	3.88	2.13
StdDev	0.76	0.64	0.89	0.71	0.92	1.07	0.99	1.36
Week 1 Off	3.6	3.1	2.7	3.2	2.9	2.8	4	3.2
StdDev	1.07	1.45	1.49	0.42	0.57	1.55	1.15	1.03
Week 1 Diff	-0.6	-0.23	1.05	0.05	0.73	1.2	-0.13	-1.08
Week 2 On	2.71	2.86	2.14	2.29	2.14	2.29	3.86	2.71
StdDev	0.95	0.9	1.07	0.95	0.9	0.95	1.07	0.95
Week 2 Off	2.79	2.57	2	2.5	2.21	1.93	3.57	3
StdDev	0.89	1.28	0.96	0.76	0.89	0.83	1.02	1.24
Week 2 Diff	-0.07	0.29	0.14	-0.21	-0.07	0.36	0.29	-0.29
Week 3 On	3.7	3.4	2.8	3	3	2.5	3.9	3
StdDev	0.48	0.7	1.03	0	0	0.71	0.74	0.67
Week 3 Off	3.31	2.77	2.77	2.92	2.85	2.85	3.69	3.15
StdDev	1.25	1.36	1.36	1.26	1.21	1.14	1.55	1.21
Week 3 Diff	0.39	0.63	0.03	0.08	0.15	-0.35	0.21	-0.15
Week 4 On	3.07	2.4	2.13	2.47	2.33	2.07	3.33	2.67
StdDev	0.7	1.06	1.06	0.83	0.9	0.8	0.98	1.18
Week 4 Off	3.67	3.92	3.08	3.33	3	3	3.83	3.25
StdDev	0.98	1	0.9	0.49	0.43	0.85	0.83	0.97
Week 4 Diff	-0.6	-1.52	-0.95	-0.87	-0.67	-0.93	-0.5	-0.58

D	Uncert	Distract	Tired	Hard	Int	Joy	Surp
Week 1 On	3.88	2.25	2.38	2.75	3	3.13	3.38
StdDev	0.64	0.46	0.92	1.04	1.07	0.83	0.92
Week 1 Off	3.4	2.3	2.9	3.4	3.2	3.4	3.8
StdDev	1.07	0.82	0.99	1.26	1.32	0.84	0.92
Week 1 Diff	0.48	-0.05	-0.53	-0.65	-0.2	-0.28	-0.43
Week 2 On	2.86	2.57	3	4	3	2.86	3.14
StdDev	1.57	1.27	0.82	0.58	0.82	1.07	0.9
Week 2 Off	3.21	2.93	3.36	4.5	3.29	2.93	3.14
StdDev	1.05	1.07	0.93	0.52	1.07	0.83	0.77
Week 2 Diff	-0.36	-0.36	-0.36	-0.5	-0.29	-0.07	0
Week 3 On	3.3	2.3	2.9	4.3	4.4	4.1	3.4
StdDev	0.82	0.95	0.74	0.67	0.84	0.99	0.84
Week 3 Off	2.62	2.69	2.69	3.62	4	3.15	3.62
StdDev	1.26	1.75	1.44	0.87	0.71	1.07	0.65
Week 3 Diff	0.68	-0.39	0.21	0.68	0.4	0.95	-0.22
Week 4 On	3.53	2.33	2.73	4.2	3.6	2.87	3.2
StdDev	0.83	0.9	0.8	0.68	1.24	1.13	0.86
Week 4 Off	3.5	2.92	2.67	4.08	3.75	3.42	3.17
StdDev	0.52	1	1.23	0.79	0.87	0.67	0.72
Week 4 Diff	0.03	-0.58	0.07	0.12	-0.15	-0.55	0.03

Table 3.3: Relationship Between Deception and Confounding Factors Week by Week

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time
Week 1 On	3	2.88	3.75	3.25	3.63	4	3.88	2.13
Week 1 Off	3.6	3.1	2.7	3.2	2.9	2.8	4	3.2
Week 2 On	2.71	2.86	2.14	2.29	2.14	2.29	3.86	2.71
Week 2 Off	2.79	2.57	2	2.5	2.21	1.93	3.57	3
Week 3 On	3.7	3.4	2.8	3	3	2.5	3.9	3
Week 3 Off	3.31	2.77	2.77	2.92	2.85	2.85	3.69	3.15
Week 4 On	3.07	2.4	2.13	2.47	2.33	2.07	3.33	2.67
Week 4 Off	3.67	3.92	3.08	3.33	3	3	3.83	3.25

D	Uncert	Distract	Tired	Hard	Int	Joy	Surp
Week 1 On	3.88	2.25	2.38	2.75	3	3.13	3.38
Week 1 Off	3.4	2.3	2.9	3.4	3.2	3.4	3.8
Week 2 On	2.86	2.57	3	4	3	2.86	3.14
Week 2 Off	3.21	2.93	3.36	4.5	3.29	2.93	3.14
Week 3 On	3.3	2.3	2.9	4.3	4.4	4.1	3.4
Week 3 Off	2.62	2.69	2.69	3.62	4	3.15	3.62
Week 4 On	3.53	2.33	2.73	4.2	3.6	2.87	3.2
Week 4 Off	3.5	2.92	2.67	4.08	3.75	3.42	3.17

Table 3.4: Magnitude of Confounding Factors Week by Week

We thus conclude that, for this sample, confounding factors had some significant correlations with type 1, type 2, and type 3 errors relative to the presence or absence of deception.

Table 3.4 summarizes the results based only on the ratings of the confounding factors week by week. When deception was enabled, perceived success became worse with time, while when deception was disabled, perceived success became greater with time. Success was always considered important, but decreased slightly in import over time. Time pressure tended to increase over time for those under deception but not for those not facing deception. The lowest uncertainty was experienced with deception on, but generally did not correlate with the presence or absence of deception. Exhaustion was not correlated with these activities. All of the efforts were considered difficult to the participants with the exception of the first week which was very easy to complete, even if it was very hard to detect the deception. Interest and enjoyment were very high in the third week.

More interesting results come in terms of difficulty, interest, enjoyment, and surprise. The first week was an extremely easy exercise designed to assure that all teams would believe they had achieved their objectives. The assessment of its ease would likely be more stark if they had the experiences in a different order, however, it is plain to see that it was easier by the results. Week 3 was considered quite difficult, most interesting, and most enjoyable by a significant amount. This particular exercise was pretty action packed in the sense that there were always things to see, things to try, and things going wrong. It was pretty interesting to watch as well. It was intentionally designed to induce errors of omission by providing massive quantities of information - more than could possibly be analyzed in the time allotted. Frustration ran high in a few instances, but clearly the participants enjoyed the effort, were engaged in the activity, and it appears that it drove them toward high-level cognitive activities.

3.10.3 Experiments 4-6 Taken as a Group

There was a one week pause between week 4 and week 5 to allow teams to improve technologies in use and rethink the previous results. While the teams did not have their full time to spend on this effort for that period, all participants were also required to study attack scripts available over the Internet, were provided a series of training sessions on "red teaming", and were provided with classes on the use of command scripts for systems administration tasks. They indicated that this had a substantial improvement in terms of their skills. At the end of week 4, teams were provided with detailed reviews of the previous experiments including brief solutions on how they could go about defeating the defenses that were in place. This included specific details on how to defeat the systems in hop 4. In addition, insider supervisor access was provided to a system within the firewall for hop 4 so that attackers had detailed information comparable to that granted to an undetected attacker in an overrun situation. The attackers were also provided with full access to copies of all of the hardware and software systems used in the experiments, all network diagrams were accurate to the level of detail provided, and after the fifth week were provided with specific training and tools that would allow them to bypass the problems they had the most problems with during the previous week.

When we plot hops 4-6 as if they were separate attacks (Figure 3.10) we see that teams acting without deception working against them tend to get further in the real attack graph faster and that the effects of learning improve performance of attackers not exposed to deception. On the other hand, attackers exposed to deception do not appear to make progress deeply into the attack graph more quickly after being exposed to the situation repeatedly. They do remember some of their successes but they also appear to remember the deceptive information and continue down deception paths despite repeated exposure, additional training, increasingly detailed information about how the deceptions work, and increased training on attack tools.

When plotted as if the attack continued week after week with a 'reset' to level '0' at the end of each session (Figure 3.11) we see that teams without deception learn to avoid self-deceptions and move through the attack graph more rapidly on subsequent tries, slowing as they reach the previous limit and moving forward slowly from there. Teams facing deception tended to give up and continued to go down deception paths even after more than 10 hours of effort. They repeatedly encountered the same deceptions and were unable to avoid them, apparently because they were unable to differentiate the deceptions from the real situation. In Figure 3.11, coloring differences are used to differentiate groups.

The Friday group decided to cease participation in the attack activity by the 5th week because it was perceived as a waste of time. In interviews with the test subjects, they indicated that they would prefer to spend their time on something more useful such as other projects that they worked on. This team (starting in week 4) consisted almost entirely of individuals who did not work every day of the week and thus had more limits on their time than other participants. It is not known whether this movement to other things represents an effect of deception, however, this team was operating with deception enabled during week 4 and made the most progress in their efforts of any team with deception enabled (they reached +3 at 3.5 hours after reaching -4 at 2.75 hours). Their forms indicated substantial frustration and high difficulty in week 4 as well as low enjoyment, all factors that we might predict would lead to resignation from the activity.

The Thursday team also decided to cease participation, but in their case this happened after the 5th week. They were also working against deceptions and they asserted that they were more interested in another student activity at that time. It is particularly noteworthy that this group decided to stop after reaching +3, just as the Friday group did. They first reached +3 at 3:15 in their second 4 hour session (or at 7:15 relative to the original start) and decided not to continue the next week.

The only remaining group working against deception (Monday) reached +4 only after 11 hours of participation and never reached +5. Both teams working without deception reached level 4 in the first 3 hours and re-attained it before the Monday group to achieve it for the first time. The

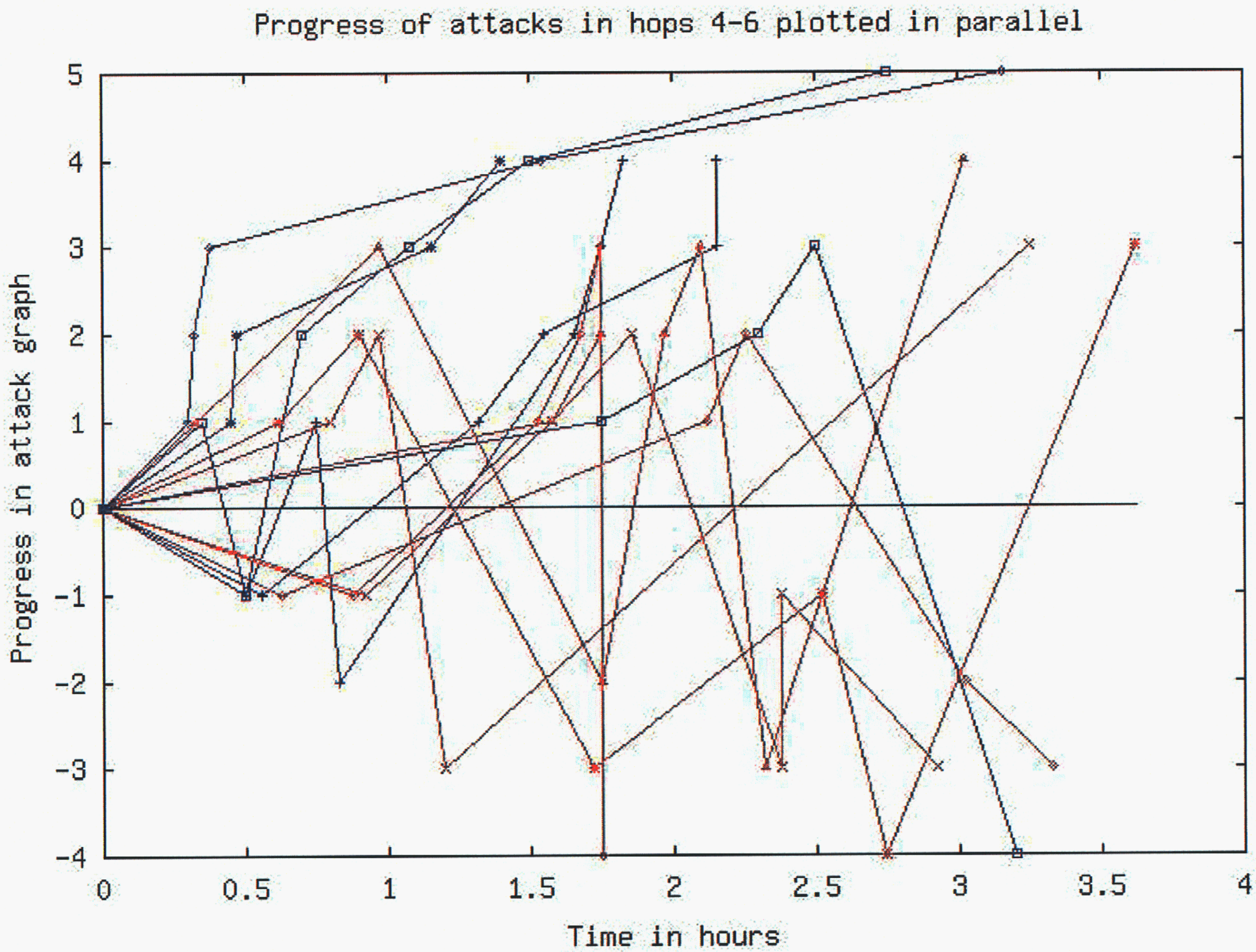


Figure 3.10: Progress of Hops 4-6 in Parallel

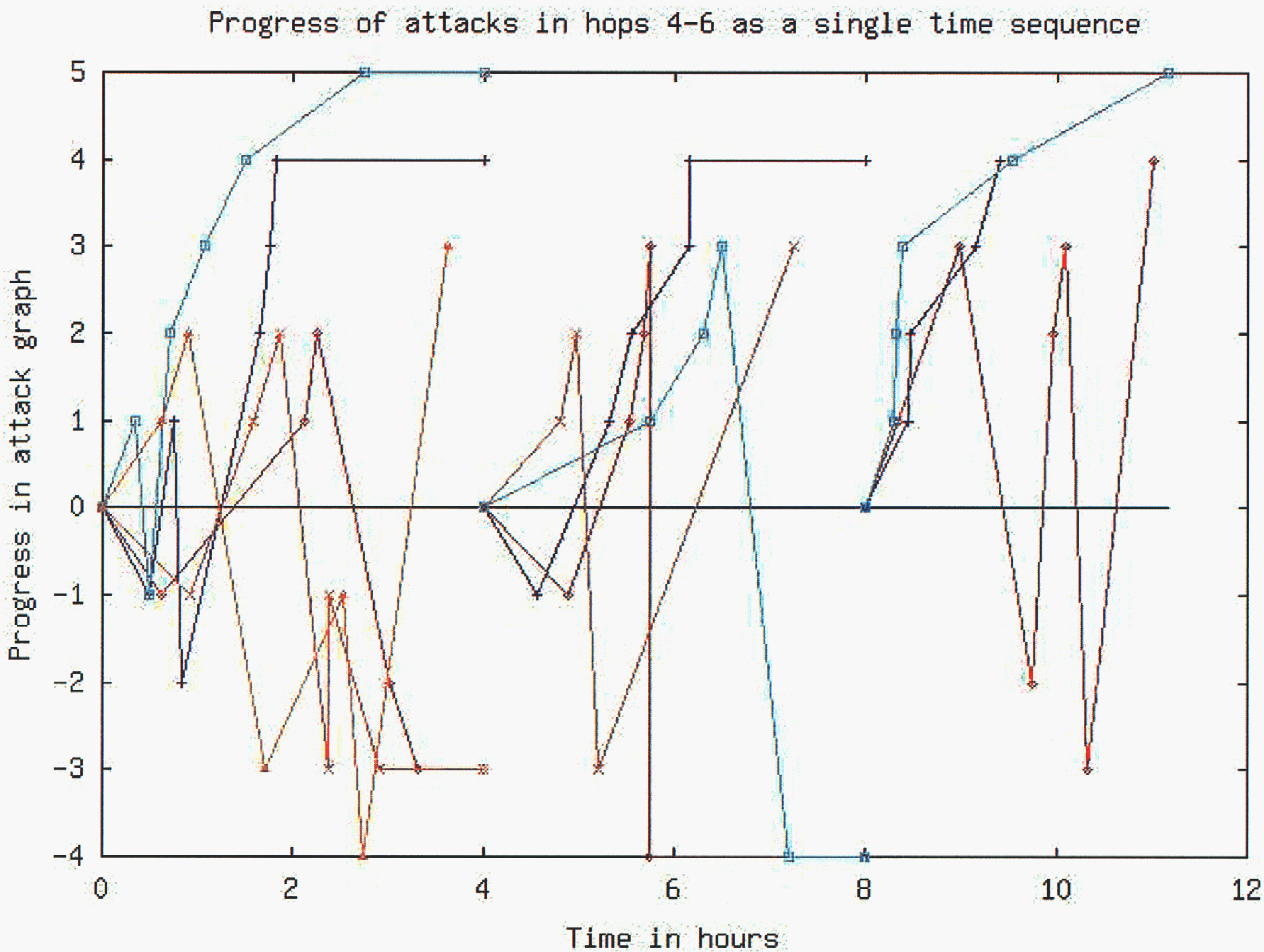


Figure 3.11: Progress of Hops 4-6 in Sequence

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time
Deception On	3.1	2.79	2.24	2.59	2.31	2	3.38	2.24
StdDev-On	0.82	1.15	1.27	0.95	0.89	0.93	0.94	1.09
Deception Off	3.61	3.48	2.9	3.13	2.87	2.84	3.87	3.13
StdDev-Off	0.93	0.82	0.76	0.63	0.48	0.91	0.86	1.06
Diff (on-off)	-0.51	-0.69	-0.66	-0.54	-0.56	-0.84	-0.49	-0.89

D	Uncert	Distract	Tired	Hard	Int	Joy	Surp
Deception On	3.55	2.79	3	4.38	3.34	2.86	3.28
StdDev-On	0.87	1.29	1	0.62	1.2	1.22	0.75
Deception Off	3.42	2.61	2.94	4.19	3.39	2.94	3.19
StdDev-Off	0.77	0.84	1.13	0.76	1.07	0.94	0.81
Diff (on-off)	0.13	0.18	0.06	0.19	-0.04	-0.07	0.08

Table 3.5: The Relationship Between Deception and Compounding Factors for Weeks 4-6

only group not undergoing deception to reach level -4 deceived itself by not ignoring its own packets in its analysis for a short period of time and recovered from this very quickly.

Deception clearly slowed the attacks, total progress against defenses is far worse when deception is present, and in this case, that attackers tend to abandon attacks in the face of deception while those not facing deception did not abandon the attacks.

3.10.4 Confounding Factors in Weeks Four to Six

We already mentioned that the Friday group abandoned the effort after confounding factors reached levels of 4/5 or above in their self-assessments. The data in Table 3.5 shows the effects of deception on the confounding factors far more clearly. It is important to note that the number of samples became quite small at the end since only 3 out of the original 15 participants continued to participate (1 in 5). For the group not encountering deception, 8 out of 12 initial participants continued through the end of the sequence.

According to this data, the confounding factors related to the cognitive effects of deception are not strongly correlated to the presence of deception, but there is a correlation in some areas. For example, while surprise, enjoyment, interest, distraction, uncertainty, and difficulty were relatively uncorrelated to the presence of deception at this point, time pressure, desire for success, and planning indicators were negatively correlated with the presence of deception on levels at or near a standard deviation. This would seem to tend to indicate that an expectation of failure built up when deception was present, resulting in lowered expectations, less trust in planning and leadership, and, interestingly, less of a feeling of time pressure. As the desire and expectations of success were reduced, time apparently became less of an issue.

Things get even more interesting as we examine the time effects of deception (Figure 3.6). Note that because a large portion of those undergoing deception opted to stop their efforts, the data values of those who did not participate are not present in the statistics when they are not participating. The removal of the participants with the least interest and enjoyment are likely the reason there is not a large negative correlation of enjoyment with deception. In exit interviews those who left indicated that they were not enjoying the activity very much and that their interest was falling off in favor of their other work. Difficulty was perceived as very high for this effort by all parties, and particularly more difficult, distracting, and uncertain in the second week for those who subsequently left. As the perception of potential for success was reduced the teams also became less able to work together.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time
Week4-Off	3.67	3.92	3.08	3.33	3	3	3.83	3.25
StdDev	0.98	1	0.9	0.49	0.43	0.85	0.83	0.97
Week5-Off	3.5	3.36	2.93	2.93	2.79	2.64	4	2.93
StdDev	0.94	0.84	0.83	0.62	0.58	1.08	0.88	1.21
Week6-Off	3.88	3.38	2.88	3.25	2.88	2.88	3.88	3.13
StdDev	0.83	0.52	0.64	0.71	0.35	0.35	0.99	1.36
Week4-On	3.07	2.4	2.13	2.47	2.33	2.07	3.33	2.67
StdDev	0.7	1.06	1.06	0.83	0.9	0.8	0.98	1.18
Week5-On	2.91	3	1.91	2.64	2.09	1.64	3.55	1.73
StdDev	0.93	1.04	1.46	1.28	0.93	0.92	0.74	0.71
Week6-On	4	4	4	3	3	3	3	2
StdDev	1	1	1	0	0	1	1	1

D	Uncert	Distract	Tired	Hard	Int	Joy	Surp
Week4-Off	3.5	2.92	2.67	4.08	3.75	3.42	3.17
StdDev	0.52	1	1.23	0.79	0.87	0.67	0.72
Week5-Off	3.5	2.86	3.29	4.36	3.14	2.71	2.86
StdDev	1.02	0.77	0.61	0.74	0.95	0.99	0.77
Week6-Off	3.25	2.13	2.75	4	3.25	2.75	3.63
StdDev	0.46	0.99	1.39	0.76	1.28	0.89	0.74
Week4-On	3.53	2.33	2.73	4.2	3.6	2.87	3.2
StdDev	0.83	0.9	0.8	0.68	1.24	1.13	0.86
Week5-On	3.64	3.64	3.55	4.64	3	2.82	3.45
StdDev	1.07	1.3	0.89	0.53	1.04	1.06	0.76
Week6-On	3.33	2	2.33	4.33	3.33	3	3
StdDev	0.58	1.73	1.15	0.58	1.53	1.73	0

Table 3.6: Weekly Deception-differentiated Compounding Factors for Weeks 4-6

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time
Week4-Off	3.67	3.92	3.08	3.33	3	3	3.83	3.25
Week5-Off	3.5	3.36	2.93	2.93	2.79	2.64	4	2.93
Week6-Off	3.88	3.38	2.88	3.25	2.88	2.88	3.88	3.13
Week4-On	3.07	2.4	2.13	2.47	2.33	2.07	3.33	2.67
Week5-On	2.91	3	1.91	2.64	2.09	1.64	3.55	1.73
Week6-On	4	4	4	3	3	3	3	2

D	Uncert	Distract	Tired	Hard	Int	Joy	Surp
Week4-Off	3.5	2.92	2.67	4.08	3.75	3.42	3.17
Week5-Off	3.5	2.86	3.29	4.36	3.14	2.71	2.86
Week6-Off	3.25	2.13	2.75	4	3.25	2.75	3.63
Week4-On	3.53	2.33	2.73	4.2	3.6	2.87	3.2
Week5-On	3.64	3.64	3.55	4.64	3	2.82	3.45
Week6-On	3.33	2	2.33	4.33	3.33	3	3

Table 3.7: Magnitude of Compounding Factors for Weeks 4-6

The clearest change in characteristics happens at the break between week 5 and week 6. The uncertainty, distraction levels, tiredness, and difficulty factors all increase in the deception-on group from week 4 to week 5, and as an apparent result, in week 6 these participants leave and levels return to more moderate levels. Similarly, in this period, interest and enjoyment drop in the group of people who left, dragging down the average, which then increases after this group leaves. Similar events appear to correlate for the group that left at the end of week 4, however, the overall effect is less stark because there are more people in the sample space at this time.

3.10.5 Special Runs

In addition to the runs described above as series of 'special runs' were carried out both as demonstrations. These did not include control groups and were under unique circumstances. The results are illuminating and are thus provided here, but the reader is cautioned that they are each special cases and conclusions are thus more speculative for now. These results are not plotted with other results or collated in summary tables because they are not statistically meaningful. They do, however, represent examples of experiments that might be repeated under controlled conditions in the future so as to provide additional statistically valid results.

SR-1 was a slight variation on experiment 3 in which the increased rotation mechanism was not operating and the Trojan horses in the real victim system were very stark and obvious. It serves as a very good example of an insider team with local access because:

- It involved two more experienced attackers as participants, one with many years of experience in testing systems for vulnerabilities and another who runs higher quality attack research and development teams,
- It included team members who had previously done exercise 3, both with deception on and deception off, and a team member who had previously supervised experiment 3 as the observer, including doing the system configuration.
- False traffic was introduced using an additional method to better simulate a real environment of this sort.
- The experiment was done after all of the other experiments and included team members who had been through the entire sequence of previous experiments.
- Several team members had been briefed on deception and deception techniques as part of the project team,
- Training and tools were provided specifically on how to detect and bypass the sorts of deception in use, and
- Specific training on how to rapidly accomplish denial of service was given, including an example demonstrated in less than 15 seconds which served as the prototype for the actual denial of service attack used.

In this case, the team was successful in 2R (in that they started to seek content), but doubted their results and never acted on them (thus 2D), and was successful in 3R but didn't fully believe it even as the experiment ended (thus 3D). All they convinced themselves of was that (1) they had found something that they did not believe to be what they were looking for and (2) that when they were running out of time, they denied services to themselves. This seems to indicate clearly that the knowledge of the possible presence of unknown deception has very beneficial defensive effects on potential attackers and it strongly supports the notion of publishing results on deception at this level of specificity.

Perhaps the most important result of this effort is that these deceptions were quite effective for a substantial period of time against non-expert attackers even though the quality of the dazzlement deceptions made them fairly easy to differentiate from live traffic. As we worked on improving this quality within fairly limited scope, we were surprised at how important the improvement in quality can be to the effectiveness of the deception.

SR-2 was a custom designed deception which was run against a set of red teams from the Naval Postgraduate School in Monterey, CA.

- In this run, a CD-ROM-based deception system was brought on site and setup by two people in just over one hour. This included wiring of about 7 computers and 5 hubs together (one computer already contained 5 Ethernet interface cards), booting from CDs, configuring interfaces, and setting up a set of browsable servers and browsing workstations.
- The experiment then ran for 3.5 hours with a red team consisting of about 8 attackers including graduate students and a professor. The team had sets of tools and a good background in this sort of activity. They were also provided with an 'insider team' of less advanced students who started on the same hub as the systems used for browsing web sites, limited message communication with the insiders, and an oracle who truthfully answered all questions.
- As time progressed, more and more information was fed to the attackers to facilitate their efforts, and eventually, they were given access to the browser hub, then the server hub, and were provided with user IDs and passwords for the systems under attack.
- The defenses in this case were extremely disruptive to unauthorized users. They presented a combination of deceptions designed to interfere with their normal operations and a high bandwidth of modified real-time traffic feed from the rest of the facility. They included type 1, type 2, and type 3 deception elements and there was time pressure.

In this experiment, deceptions were specifically directed toward tools that analyze networks. The intent was that these tools would be rendered ineffective through the use of protocol-level deceptions that caused them to interpret information incorrectly and give the user false information as a result.

The deceptions in this case were very effective. The attack tools tended to provide inconsistent information to the user which could not be sensibly analyzed and understood in context. In addition, the tools used by the attackers to defend themselves were sounding alarms because they misinterpreted the protocol deceptions as attacks against their computers. In effect, the attackers fooled their own detection systems into believing that they were under attack.

After a point the attackers decided that additional intelligence information would be of no use. It simply didn't matter what they were told, the information was of no value because their tools were incapable of exploiting the knowledge. For example, their tools provided no means to selectively ignore information that caused them to be unable to connect to remote computers.

They also attempted denial of services attacks, but ended up denying their own services by accident. They found that they could not even communicate between themselves over their computers because the deceptions were so disruptive to unauthorized traffic that their communications tools failed. Attempts to collect statistical data were also disrupted by pseudo-random response changes so that any attempt at gaining consistent information was fruitless.

During the entire activity 'legitimate' users were continuing to access the web servers from their browsers unaffected. Because of the substantial volume of altered legitimate traffic in the network, the attackers were unable to home in on which traffic came from real browsers and servers within the time available.

The most similar previous experiment was experiment 3, but in this case the deception defenses were significantly harsher and more aggressive than in previous runs. Using the same metrics as used in experiment 3, these attackers reached 1D within a few minutes of the start of the effort and never got any further.

3.11 Group Behavior Under Deception

In experiments reported in this paper we have access to information about (1) what the individuals in the group said and did, (2) how the group interacted, (3) survey results, and (4) expert notes on technical progress. In many real-world deceptions, only the remote reflection of the actions of an intruder might be available for analysis. Making the connection between these remote reflections and their root causes would be helpful in adapting the presentation of deceptions to targets so as to increase their efficacy and drive attackers and groups of attackers toward specific cognitive situations.

For example, in some cases, we might wish to keep the intruder interested in the deception target so that their location can be traced, their 'hand' identified, and their methods observed; while in other cases we might want to encourage the intruder to move on. The behavioral approach discussed below has the potential to provide the information required to use these remote reflections of behavior in this manner.

3.11.1 A More Detailed Examination

We focused on one group for more in-depth analysis. The group chosen was composed of individuals who were fairly well matched in skill levels and about average for the students participating in the entire series of exercises. The group was notable for their participatory and democratic decision style. They were an experimental group, where deception might be turned on or off based on chance, their two previous exercises were without deception. The session analyzed here was the group's first experience with deception enabled, and was the third in the series (Hop 3). They were familiar with the exercise routine, but had not yet been briefed on the design and motivations for the experiments. Other details are as described earlier in this paper.

The group's work partitioned into two parts. The group achieved 1R and 1D at about 50 minutes into the first hour of the four-hour exercise. Observing behavior in this segment gave the impression of steady, purposeful activity. In the remainder of the exercise, they made no further progress in the attack graph.

3.11.2 Analysis Methods

In our analysis we considered only behavior, the duration of an action by one or more members of the group, without attending to the cognitive content of, or motivation for, the action. Behavioral responses were scored using a video player with a counter (approximately one count for every two seconds), recording the results directly into an Excel spreadsheet. The scoring indicated the counter value (time), the actor initiating a response (m, c, j, gs and/or g for the group as a whole), and to whom the response was directed, which could include the actor (self), other actors, or a computer. All scores were recorded as a one-directional map from actor(s) to receiver(s). The group or individual orientations were noted with each action (e.g., "facing computer", "sitting in a circle", "sitting in a semicircle watching j's monitor"). Finally, the action was succinctly described, (e.g., "typing into the computer", "watching the monitor", "commenting on c's action"). This method of scoring can be generalized in future experiments to distributed groups interacting only via computer keyboard input to a network, or in comparison experiments evaluating individuals working alone, face-to-face groups, and distributed groups.

For the subset of data presented here, we counted the duration of response as the time between initiation of one response and the start of the next. Response duration was chosen because of its simplicity; only one-directional behavior needs to be tracked; reciprocal response of the receiver depends on the actor, and is not independent. Scoring of response duration also lends itself to automation.

For the group (all actors) evaluation, a response could be by any one or more members of the group. The scoring transcript was also divided into secondary transcripts for each individual actor

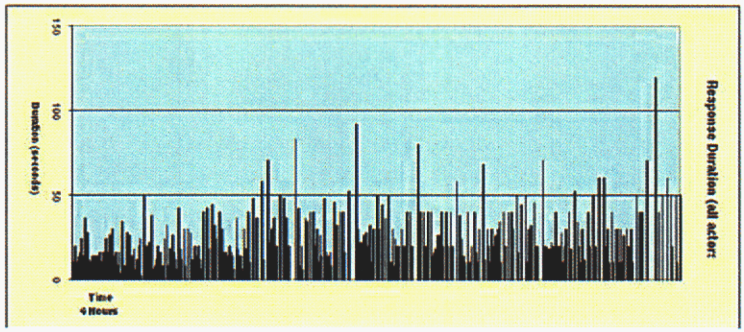


Figure 3.12: Frequency of Group Response

to discern if individual patterns differed significantly from each other and from the group as a whole. For the evaluation of response duration for individual actors, the scoring represented the duration between one action by that actor, and the next by the same actor. The response duration for the "all actors" tables and charts are not strictly commensurable with the individual actor tables and charts. In fact, the scoring approximately doubles the number of responses for the individual tallies, and response durations (i.e., time till next response) can be much longer. The individual actor data bore out the group's democratic work style showing remarkably similar patterns for each. This data is available in appendixes of to this paper.

3.11.3 Limits on the Method

The counter on the video seemed to skip values, possibly during pauses of the tape. We spent a lot of time with a stop watch to see if calibration of the tape was accurate, measuring frequently throughout the scoring, and it was accurate for all time segments measured, yet the values of the full count came out about one hour short in a ten hour scoring session. As far as we can tell, the slippage was randomly distributed, and did not appreciably change the response duration pattern that is our focus. During the second hour of the experiment, the network had to be rebooted. Except for a slight pause in the work, this caused no apparent disturbance to the group. About an hour into the experiment a pizza was delivered to the group and did not noticeably disturb the work. The scoring information was recorded into an Excel spreadsheet and calculations and partitions of that data were checked and tallied in various ways to ensure accuracy. Transcription errors were estimated to be approximately 1%, based on the crosschecks. As far as we can tell, the errors are randomly distributed, and do not appreciably affect outcomes reported.

3.11.4 Analysis Performed

Figure 3.12 presents the frequency of group response durations, graphing data from the group behavior response analysis for the entire four-hour session. The response duration is calculated as the time lapsed from one action to the next by any actor, or the entire group. The frequency of actions was higher, and the duration shorter for the first hour when the group worked purposefully to the achievement of the first milestone in the attack graph. This is seen as more lines per horizontal distance and lower height of those lines. The frequency of actions decreased and the duration increased as the group spent more time observing network traffic and trying to discern some pattern to it. This is reflected in increased height of lines (response durations) and decreased frequency of lines toward the right of the chart. This is associated with watching traffic, uncertainty about what actions to take, and observing effects for those tentative actions the group did take.

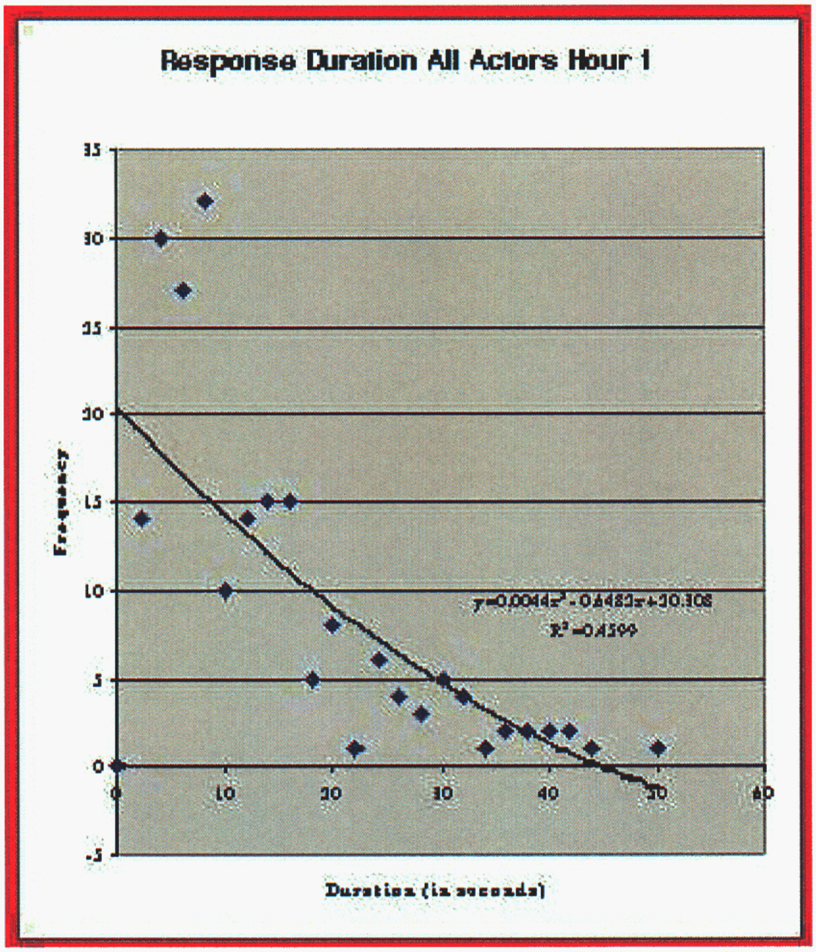


Figure 3.13: Group Responses from Hour 1

3.11.5 Analysis Performed

The change from purposeful to less coordinated behavior is captured in two scatter plots showing the correlation between frequency and duration (in seconds) of responses. Figure 3.13 is based on data from the first hour for the whole group. There is a moderate negative correlation between frequency and duration of responses; the distribution is skewed to the left; and the standard deviation for this data is approximately half that for Chart 3, which is the data set representing the last three hours of the exercise. This indicates lower variability.

Figure 3.14 displays the frequency vs. duration scatter plot for hours 2-4 of the same experiment. The random, disoriented searching activity after the first hour is reflected in the virtually flat regression line fitted to the scatter plot data. The frequencies of response duration of the group are not correlated.

3.11.6 Conclusion

The results of behavioral analysis of the group session demonstrates that a simple scoring method and scatter plots for correlation between frequency and duration can give useful pattern information about actions of individuals and groups attacking computer networks. This may also be a helpful

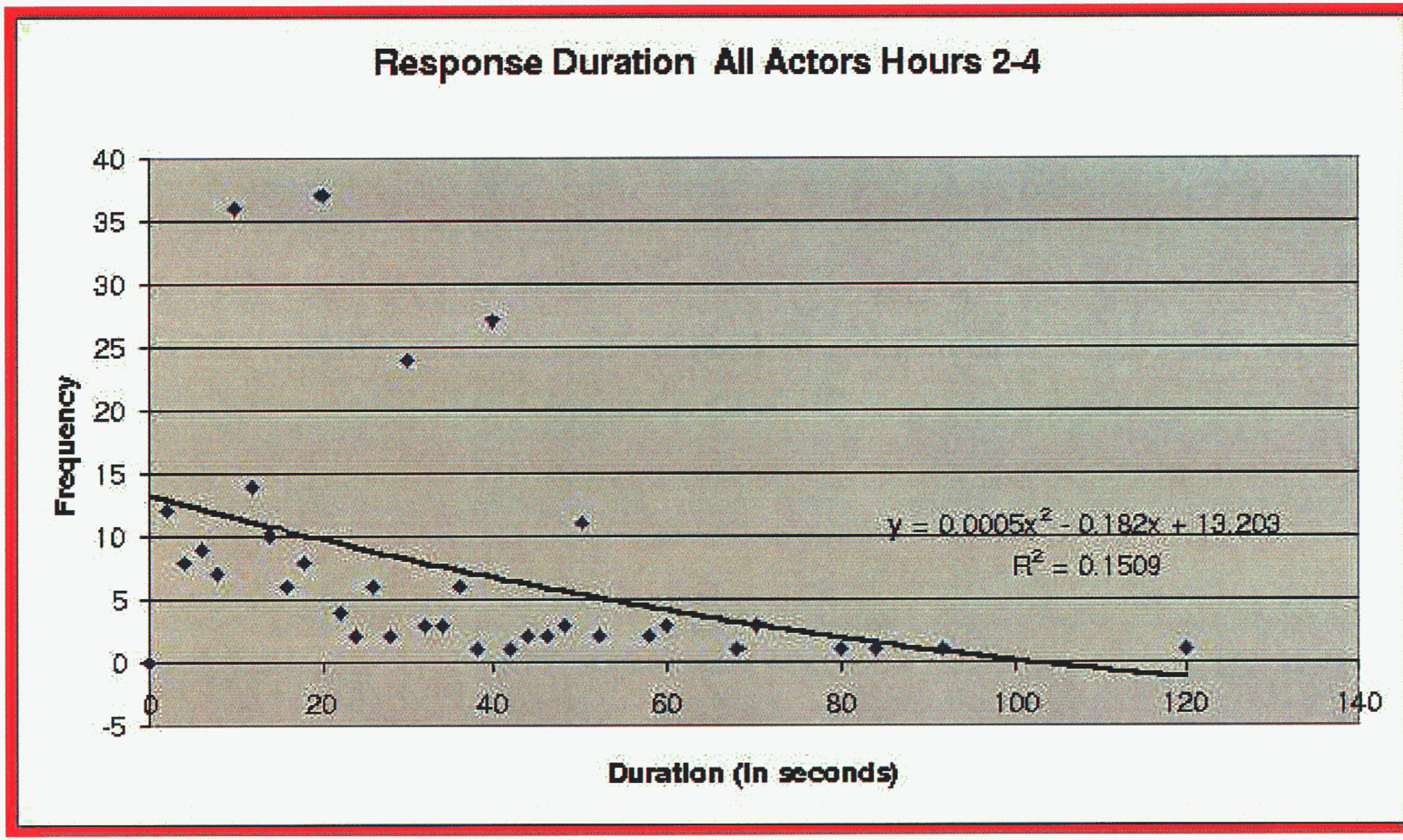


Figure 3.14: Group Responses from Hours 2-4

adjunct in monitoring tunable deceptions.

3.12 Summary, Conclusions, and Further Work

Based on these results it appears that the network technology deception capabilities are very effective at what they do, but that in order to be far more convincing for a far longer time against more skilled attackers, it will be necessary to create improved content-oriented deceptions.

These results clearly show that psychological factors identified elsewhere are key components of the effects of deception on computer network attack groups. They demonstrate that highly effective deceptions can be implemented, that attackers fail to detect them when we induce type 3 errors, that they fail to differentiate them when we introduce type 1 errors, that they give exceeding weight to any information when we induce type 2 errors, and that the whole set of factors investigated in earlier works plays into the effectiveness of deception in defending information systems. Attackers showed backtracking behaviors, group cohesion and performance was highly affected by deceptions, and issues like group think clearly came into play in causing groups to make worse decisions than individuals might have otherwise made.

The attack graph methodology of measuring progress over time seems to be very useful in this effort and it appears to be a good methodology for understanding progress of attacks and efficacy of defenses over time. The use of a marking scheme for tracking group movements also shows promise and seems to lead to the potential for associating detectable behavior with group activities and behavior.

The net effect of deceptions is that attackers spend more time going down deception paths rather than real paths, that the deception paths are increasingly indifferentiable to the attackers, and that the defenders can gain time, insight, data, and control over the attackers while reducing defensive costs and improving outcomes. Attackers become frustrated, ineffective, and show increased attrition.

Experiments now being planned include a sequence in which 120 university students attack a set of systems over a full semester as part of classroom activity and a preliminary sequence in which a smaller collection of high school students wring out the same technology.

Finally, a pitch for more research funding. At this point we are unable to do an adequate job of analysis necessary to complete a model that would link observed individual and group behavior to observables from a defender's perspective. This is critical to making adaptive defenses. Similarly, our ability to analyze data and configure new experiments is quite limited. While we have been able to leverage other funds into the creation of testbeds for experiments, we have no funds for running experiments or performing analysis of results. As a result, the upcoming experiments will operate but we will likely be unable to analyze results in an in-depth manner and theoretical progress will be unlikely to proceed much further than it has to this point.

We urge other researchers to pick up where we have had to leave off. While we cannot provide all of our test data for reasons of privacy, we are able to provide significant anonymized data to legitimate researchers and we welcome anyone who is interested in doing further statistical analysis or other characterizations to work with us on this effort.

Chapter 4

Leading Attackers Through Attack Graphs with Deceptions

by Fred Cohen and Deanna Koike[1]
May 29, 2002

- Fred Cohen: Sandia National Laboratories
- Deanna Koike: Sandia National Laboratories (CCD), UC Davis

4.1 Abstract

This paper describes a series of experiments in which specific deceptions were created in order to induce red teams attacking computer networks to attack network elements in sequence. It demonstrates the ability to control the path of an attacker through the use of deceptions and allows us to associate metrics with paths and their traversal.

4.2 Background and Introduction

A fairly complete review of the history of deception in this context was recently undertaken, and the reader is referred to Chapter 2 for more details on the background of this area. Experimental results were also recently published and the reader is referred to Chapter 3 for further details of that effort.

One of the key elements in associating metrics with experimental outcomes in our previous papers was the use of attack graphs and time to show differences between attackers acting in the presence and absence of deceptions. After running a substantial number of these experiments we were able to show that deception is effective, but little more was explored about the nature of the attack processes and how they are impacted by specific deceptions. One of the things we noticed in these experiments was that patterns seemed to arise in the paths through attacks. While this has long been described in literature that seeks to associate metrics for the design of layered defenses, and in the physical reality it has long been used to drive prey into kill zones, to date we have not seen examples of the design of such defenses so as to lead attackers into desired paths in the information arena.

[1]This chapter is published online at `http://all.net/journal/deception/Agraph/Agraph.html` and also appears in [CR03].

Our ongoing theoretical work led us to the notion that in addition to measuring paths through attack graphs over time, we should also be able to design attack graphs so that they would be explored in a particular sequence. By inducing exploration sequences, we should then be able to drive the attackers into desired systems and content within those systems. Indeed, if we become good enough at this, we might be able to hold attackers off for specified time periods by specific techniques, change tactics automatically as attackers explore the space, so as to continue to drive them away from actual targets, and otherwise exploit the knowledge for both deception and counterdeception.

In this paper, we describe a set of experiments in which we used a generic attack graph and specific available techniques to design sets of deceptions and system configurations designed to lead attackers through desired paths in our attack graph.

4.3 The Attack Graph

Based on previous work already cited, we developed the generic attack graph shown in Figure 4.1, which is intended to describe, at a specific level of granularity, the processes an attacker might use in attacking a computer system.

The process begins at 'Start' and is divided into a set of 'levels' which we can number as -4 through 4 inclusive. The attacker starts at level 0 and generally moves toward increasingly negative numerical values as they are taken into a deception and increasingly higher numerical values as they succeed at attacking real victims. Lines with arrows represent transitions and each node in the graph represents a complex process which we have not yet fully come to understand. There are a lot of transitions that cross multiple levels of the graph. For example, an attacker in a real system can be led into a deception by 'tripping across' a deception within that system that deflects the attack into a deception. In addition, there is a general 'warp' that extends throughout the graph in the sense that from any given state, it is possible to leap directly to another state, however this appears to be fairly low probability and has not been well characterized yet.

Two processes are defined here, one starting with a systematic exploration of the target space and the other through random guessing. We have sought out other strategies to depict, but have found none. It appears that transitions in this attack graph are associated with cognitive processes in the groups, individuals, and systems used in the attack process as they observe, orient, decide, and act on signals they get from their environment.

4.4 Our Experimental Design

Early in 2002, we created a series of experiments in which we attempted to design sets of interacting deception-based defenses with the objective of inducing attackers to follow specific paths through the generic attack graph. For example, in our first experiment, we decided to try to induce attackers to (1) seek targets, (2) fail to find real targets, (3) find false targets, (4) attempt to differentiate false targets from real ones, (5) seek other targets, (6) find false targets, (7) differentiate them from other false targets, (8) decide to seek vulnerabilities, (9) try to enter, (10) fail to find vulnerabilities, (11) fail to enter, (12) eventually succeed in gaining limited entry, (13) attempt to exploit access, (14) decide to try to expand access, and (15) continue the process over a period of 4 hours. We will use these numbers in the following paragraphs to associate our mechanisms with the actions we sought to induce.

Our planning process consisted of creating sets of possible targets of attack with characteristics that could be identified and differentiated with different levels of effort using available tools and known techniques. This process was driven by the 'assignment' of the team (1) which was to find user systems and try to gain specific information about a criminal conspiracy from those systems. By making the more easily identified targets more obviously false, we were able to induce the behaviors associated with the loop in which attackers (3) find false targets, (4) differentiate them as false, (5)

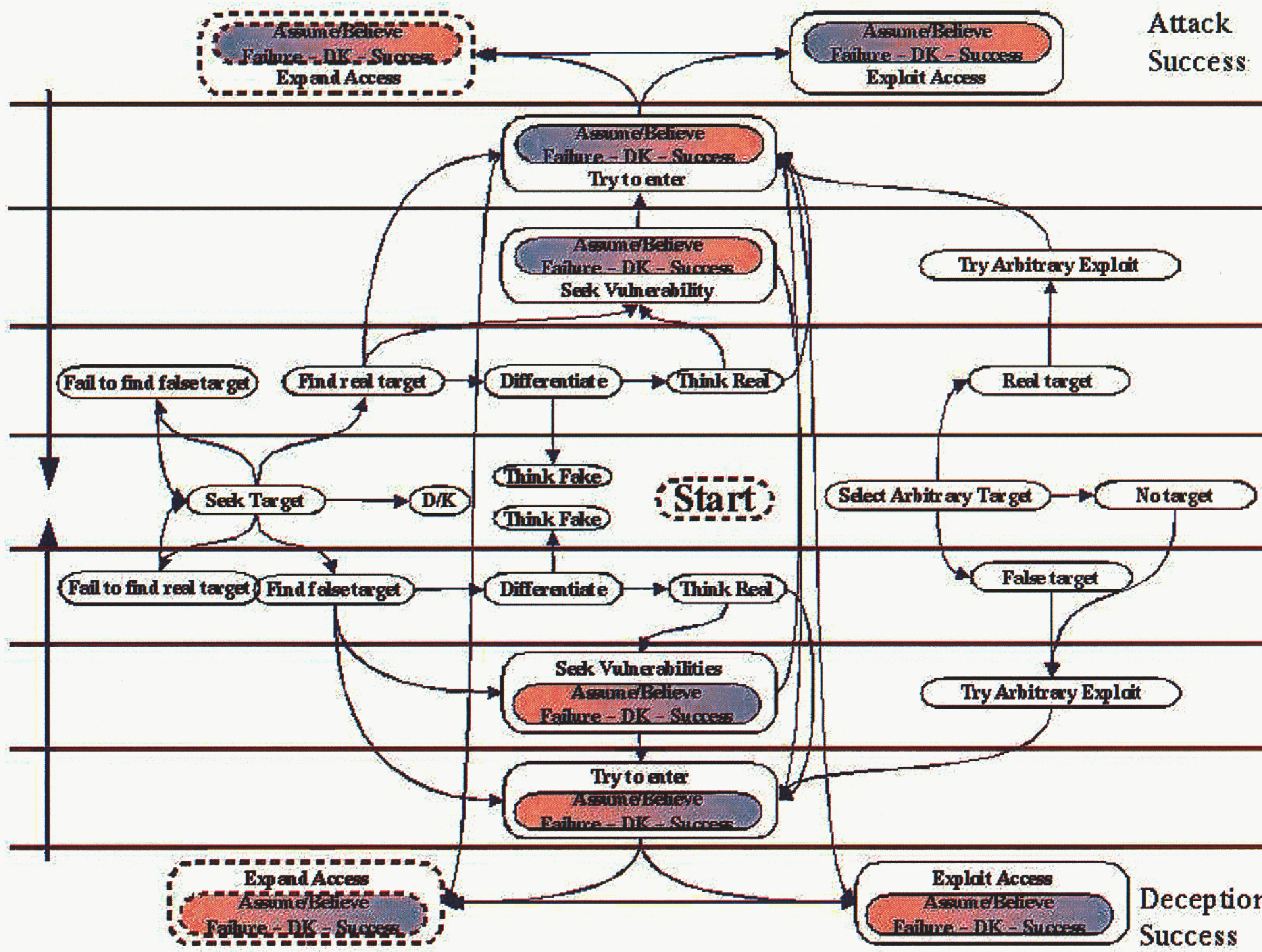


Figure 4.1: Generic Attack Graph

and seek other targets. Similarly, we used (2) concealment techniques to make it difficult to find real targets so that the attackers would be far more likely to miss them and find false targets.

To get attackers to proceed to seek vulnerabilities and try to gain entry, (6) we created real systems that were apparently in use based on normal observations. For example, (7) these systems appeared to generate traffic that would commonly be associated with users performing activity, (8) they apparently had services running on them, they appeared to respond to various probes, and so forth. The goal was for the attackers to become adequately convinced that they were legitimate targets to (9) try to gain entry. After (11) some number of failed entry attempts, (12) relatively simple entry paths were found that allowed rapid entry through apparent misconfigurations, and (13) select content implying the need for more access to get to more important content was placed in those computers to (14) entice the attackers to try to escalate privileges under the belief that this might gain them the information they sought. Some of the information that could only be obtained under escalated privileges made it very clear that this system was not the real target, thus driving the attacker back to the target acquisition phase. In addition, IP addresses were changed every few minutes and user access was terminated periodically to cause the attacker to return to the target acquisition process and attempted entry process respectively. It was anticipated that over time, these targets would be identified as false and that other targets would be sought. (15) Other less obvious targets were provided in a similar vein for more in-depth examination. Specific methods associated with these processes are described in Chapter 5. We also note that the deceptions in these experiments are fully automatic and largely static in that the same input sequence from the attacker triggers the same response mechanism in the deception system throughout the experiment.

In the first experiment, the systems being defended were on the same network as the attackers and were configured to ignore packets emitted from unauthorized IP addresses. Forged responses to ARP requests were used on all IP addresses not otherwise in use (2) to prevent ARP information from revealing real targets and ICMP responses were suppressed to prevent their use for identification of real targets.

Subsequent experiments were carried out with variations on these design principles. Specifically, we created situations in which we controlled available information so as to limit the decision processes of attackers. When we wished to hide things, we made them look like the rest of the seemingly all false environment, and when we wished to reveal things, we made them stand out by making them differentiable in different ways.

Unfortunately, we did not have the resources necessary to carry out a full fledged study in which we used the presence and absence of deception or more and less well planned deceptions in order to differentiate specific effects and associate statistically meaningful metrics with our outcomes. We did not even strictly speaking have the resources for creating repeatable experiments. Unlike our earlier experiments described in Chapter 3, in which we ran 5 rounds of each experiment with deception enabled and disabled, we had only one in-house group of attackers available to us, and of course they are tainted by each experience.

As an alternative, we created a series of experiments in which our in-house attack team was guided, unbeknownst to them, and with increasing accuracy, through a planned attack graph. We then carried out an experiment at a conference in which attack groups were solicited to win prizes (up to $10,000) for defeating defenses. The specific deception defenses were intended to induce the attackers to take a particular path through the attack graph. All attack groups acted simultaneously and in competition with each other to try to win prizes by breaking into systems and attaining various goals. No repetitions were possible, and a trained observer who knew what was real and what was deception followed the attacker activities and measured their progress.

4.5 Experimental Methodology

In each case the experiment began with a planning session in which defense team members designed a set of specific deceptions and predicted sequences of steps in the attack graph that they believed

Number	Node name	Level
0	Start	0
1	Seek Target	0
2	Fail to find false target	-1
3	Find false target	-1
4	Differentiate (Fake)	-1
5	Think Fake (from 4)	0
6	Think Real (Fake)	-1
7	Seek Vulnerabilities (Fake)	-2
8	Try to enter (Fake)	-3
9	Exploit Access (Fake)	-4
10	Expand Access (Fake)	-4
11	Find Real Target	1
12	Differentiate (Real)	1
13	Fail to find false target (Real)	1
14	Don't Know	0
15	Think Real (Real)	1
16	Seek Vulnerability (Real)	2
17	Try to Enter (Real)	3
18	Think Fake (Real)	0
20	Exploit Access (Real)	4
21	Expand Access (Real)	4
30	Select Arbitrary Target	0
31	No Target	0
32	False Target	-1
33	Try Arbitrary Exploit	-2
34	Real Target	1
35	Try Arbitrary Exploit	2

Table 4.1: Attack graph numbering

attackers would take in attempting to attack real targets. The configuration was documented and implemented and the attack sequences were discussed and put into written form as a series of states and transitions in the attack graph depicted. Numbers were associated with attack graph locations for convenience of abbreviation. These locations in the attack graph can also be roughly associated with the levels used in our previous experiments on deception. The numerical values are shown in Table 4.1.

A predicted outcome would be in the form of sequences of node numbers with a note on transitions and loops indicating the anticipated event. For example the first run starts as shown in Table 4.2.

The experiment was run with one of the defense team members taking notes on the sequence of events in terms of the attack graphs, identifying associated times. It was necessary for this team member to know specifically which targets were deceptions and which were real in order to accurately identify the location in the attack graph. With the combination of knowledge of the attack graph, the configuration, and background on deception, it is relatively easy to guess what paths are likely to occur under which deceptions. For this reason it was impossible to have the observer not know what predictions were made. Observers were trained in not revealing information about the situation to the attackers, however, this is a less than ideal situation. This represents a yet unresolved experimental limitation that can easily produce erroneous results because of the lack of

Sequence	Comment
0	Start the run
1	Seek target per assignment
2	Fail to find target (missed topology due to concealment)
3	Find false target via open ports
4,5,1	Obvious dazzlements
1,3,4,6	Limited dazzlements easily differentiated

Table 4.2: Example Predictions

Sequence	Comment
0	Start the run
1,2,1	Seek target per assignment, fail to find, return to seek
1,3,4,5,1	Find false target via open ports, obvious dazzlements, search on
1,3,4,6	Find false target, limited dazzlements easily differentiated
6,7 or 6,8,7	Obvious things to try don't work
7,8 loop	Apparent vulnerability - weak services - some not vulnerable
7,8,9	Locatable vulnerability gives user access with obvious content
9,10	Obvious content not relevant - less obvious apparent but requires privilege
9,8 or 10,8	Internal kill response mechanism kicks user out
6-10,1 or 3-4,1	Rotating IP addresses force search to restart

Table 4.3: Experiment 1 - Predictions

an unbiased observer. Note that the model implicitly assumes that at any time an attacker can revert to a previous state and that there is a low probability that an arbitrary state transition (a warp) can occur at any time from any location to another. Attack sequence prediction implicitly assumes this sort of backtracking is always possible and it is not noted unless it is specifically driven as part of the experiment.

To help compensate for this, we introduced two additional controls. During experiments, we videotape the process so that it can be independently reviewed. After the sequence of experiments, we review results with those who participated and ask them for their views of whether our depictions were accurate.

4.6 Experiment 1

In experiment 1, the predictions in Table 4.3 were made and documented prior to the start of the experiment.

Experiment 1 proceeded on 2002-04-12 using a team of 7 university students specializing in computer science and part of the Sandia College Cyber Defenders program at Sandia National Laboratories. These students all have high academic credentials, range from Sophomore to Graduate students, and have limited experience in computer attack but substantial training and experience in defending computers against attacks. They all had several weeks of previous experience with similar deception defense techniques, practice with available tools, and experience in the experimental environment.

Table 4.4 shows the results that were observed (all times relative to experiment start time). By comparing sequences we can readily see that the predicted sequences occur frequently and that there are no radical departures from the paths laid out in the prediction. The summary of predicted and

Time	Sequence	Comment
0		Configure network
0:01	1	Passive network sniffing
0:04	1,2 1,13	Designers forget 13 is present when seeking targets
0:40	1,3,4,6	Probe with broadcast ping > consistent ARPs lead to deception box
0:45	6,8	Believe it is a RedHat Linux box, try simple entry, give up too soon
0:46	8,1	Rotating IP addresses force search to restart
0:46	1,3,4.6	Find apparent real (False) target and differentiate rapidly
0:46	6.8.7	Try obvious remote root password, fail, seek vulnerabilities
0:50	7,5,4	Express that this could be a fake box - continue plan
1:01	4,7	See IP addresses changing, associate ssh service
1:20	7,1,2	After group discussion, decide to try other search methods, fail
1:30	8	Using previous results, try to access false target
1:43	8,7	Try other services
1:44	7,8 loop	Try various guesses, seek exploits
1:46	8,9	Guess valid password, gain access, see simple false content, read, believe need an exploit to escalate privileges
1:50	9,8	Internal kill response mechanism kicks user out
1:53	8,9	Regain access, identify system more closely, seek exploit
1:55	9,1	Rotating IP addresses force search to restart
1:55	1,3,4.6	Find apparent real (False) target and differentiate rapidly
2:05	6,7	Convinced they need to find ways to escalate privileges, meet to discuss
2:15	1,3,4,6 loop	Identify pattern of IP address changes for prediction of next IP address
2:30	6,7 loop	Seeking remote root attacks on fake box
2:40	6,7 loop	Notice password file (considered but did not try to crack it)
3:10	7,8	Run known remote attack, failed to work
3:15	END	Terminated for end of allotted time

Table 4.4: Experiment 1 - Observed Behaviors

Predicted	Observed	Comment
0	yes	Obvious
1,2,1	1,2,1,13	Designers forgot 13 would be present - otherwise correct
1,3,4,5,1	no	Attackers were so caught up in 1,3,4,6 that they never returned
1,3,4,6	1,3,4,6	at 0:46, 1:55, 2:15
6,7 or 6,8,7	6,8 6,8,7, 6,7	at 0:45 0:46, 2:05, 2:30 (loop)
7,8 loop	7,8 7,8 loop	at 1:44, 3:10
7,8,9	8,9	after 7,8 loops at 1:46 (missed 8,9 loop implied below)
9,10	no	never reached
9,8 or 10,8	9,8	at 1:50
6-10,1 or 3-4,1	9,1 7,1 8,1	at 0:46, 1:20, 1:55, 2:15
————	——-	————————————-
	1,13	Designers forgot to indicate real targets would be missed
	8,1	Implicit in all graphs
	7,5,4,7	Never anticipated this path (0:50-1:01)
	2,8	Use of previous results for 'direct' jump - part of other sequence

Table 4.5: Experiment 1 - Results

non-predicted sequences in Table 4.5 clarifies this comparison.

The design seems to have worked as intended, driving attackers through specific sequences of attack methods and patterns of attack. For example, there were no instances of unanticipated motions from deception to real targets, no cases in which the attackers found real targets instead of deceptions, and despite the understanding of the potential for deception by the attackers, there were no strong efforts to seek out new methods to detect other systems as long as the mysteries of the already identified systems were still being unraveled. The paths described by the attack graph were followed as if they were well worn grooves in the attackers' methods. We also note that the deception was highly effective in that the attackers never moved toward the positive 'levels' of the attack graph.

4.7 Experiment 2

In experiment 2, a more complex scenario was presented involving three networks. The attackers could move from the more distant network to an apparently more closely located network, to an inside network with the provision that once they had moved inward, they would be considered as having given up at the more distant location. There is not a lot of impetus to remain on the outside in this experiment. This then translates into three somewhat different but interrelated experiments. Each of the three experimental situations was predicted, as shown in Table 4.6, Table 4.7 and Table 4.8.

Experiment 2 used the same attackers as from experiment 1, but they were required to split into two teams and work in parallel in the same room. The measured results are shown in Table 4.9, and the results are summarized in Table 4.10.

It appears that time pressure prevented many of the potential predicted paths from being explored in this example. The exercise was just too complex for the time available. While we don't yet have a good model for the time associated with detecting and defeating various deceptions, it seems clear that the time factor played a major role in this exercise.

Sequence	Comment
0	Start - from the outside only active searches will operate
1,2 loop	Searches will fail to find real targets
1,13 loop	Searches will often fail to find false targets
1,3,4,5 loop	Some targets will be declared deceptions
1,3,4,6	A lot of seemingly different false targets will be found, some explored
6,7 loop	Attempted remote exploitation may be tried - unlikely to work
6,8 loop	Attempted direct entry may be tried very briefly (guest, guest works on some fakes)
8,9	If they gain entry, they will see obvious content
*,1	Lots of returns to 1 because of IP rotation mechanisms
	Likely to move to DMZ or Inside soon

Table 4.6: Experiment 2 - "Outside" Predictions

Sequence	Comment
0	Start - from DMZ passive observation will show content
1,2 loop	Often fail to find real targets
1,13 loop	Often fail to find false targets
1,3,4,5 loop	Some targets will be declared deceptions, much traffic will be dismissed
1,3,4,6	A lot of seemingly different false targets will be found, some explored
6,7 loop	Attempted remote exploitation may be tried - unlikely to work
6,8 loop	Attempted direct entry will be tried (guest, guest works on some fakes)
7,8 loop	Try to find other vulnerabilities and exploit them
8,9	If they gain entry, they will see obvious content
9.10	If they gain entry, they may try to autoguess the root password
8,10	Slim chance they will exploit access to stop IP rotations, defeat deception
*,1	Lots of returns to 1 because of IP rotation mechanisms
	Likely to move to Inside soon, perhaps correlate results

Table 4.7: Experiment 2 - "DMZ" Predictions

Sequence	Comment
0	Start
1,2 loop	Often fail to find real targets
1,13 loop	Often fail to find false targets
1,3,4,5 loop	Some targets will be declared deceptions, much traffic will be dismissed
1,3,4,6	A lot of seemingly different false targets will be found, some explored
6,7 loop	Attempted remote exploitation may be tried - unlikely to work
6,8 loop	Attempted direct entry will be tried (guest, guest works on some fakes)
7,8 loop	Try to find other vulnerabilities and exploit them
8,9	If they gain entry, they will see obvious content
9.10	If they gain entry, they may try to autoguess the root password
8,10	Slim chance they will exploit access to stop IP rotations, defeat deception
1,11,12,18	Some real target information may be found and dismissed
1,11,12,15	Some real target information may be found and thought real
15,16 loop	Real targets may be scanned to find possible entry points
15,17	Simple direct entry attempts may be made, denial of service attempts may be made
16,17	Scanned services will yield complex bypass mechanisms, may be bypassed
15,20	Sniffed content may be accumulated to achieve a goal
30,34,35 loop	Denial of service attempts against the network in general may be tried
30,31,32,33 loop	Denial of service attempts against the network in general may be tried
*,1	Lots of returns to 1 because of IP rotation mechanisms and real target concealment

Table 4.8: Experiment 2 - "Inside" Predictions

Time	Team 1 Seq	Team 2 Seq	Comment
0	0	0	Start - Configure networks
0:24	1,2 1,13 loop		Ping sweep of network, scripts
	1,2,4		Scripts seem to fail - verify tools by testing
0:30		1,2 1,13 loop	Arpwatch and ethereal
0:36	1,2 loop	1,2 loop	arping, nmap - yield no data
0:49		1,3,4,5	traffic from other group seen and reconciled
1:03	1,3,4,6	1,3,4,6	observe deception traffic and examine
1:15		==> DMZ	team 2 decides to move to DMZ network
1:20		1,3,4,5	dazzled responses to ethereal, arping, ping, tcp-dump
		1,3,4,6	testing det select returns (traffic actually from team 1)
1:27	==> DMZ		team 1 decides to move to DMZ network
1:27		confusion	traffic ceases (other team moved) and confusion occurs
1:30		1,3,4,5	active probing sees fake traffic - actually themselves
1:34	1,3,4,5		seeing lots of content in ethereal - result of team 2's scans
1:39	1,3,4,5 loop	1,3,4,5 loop	dazzlement of each by themselves and others
1:42	30,31,33		try arping flood - no reason - no result
1:49	1,3,4,5 loop	1,3,4,5 loop	"nmap useless"
2:00	==> Inside		team 1 decides to move to inside network
2:14		1,2 loop	confusion by not getting 'fakes' anymore (team 1 gone)
2:29	1,11,12,15		observe real traffic - unsure of what it is
2:47		1,3,4,5 loop	found self as only system in network
2:47		==> Inside	team 2 decides to move to Inside network
2:51	1,3,4,6,8,5,1		try to ssh to every IP, mirror ssh to self, return to start
2:56		1,3,4,5	confirmed mirroring behavior, noticed strange packet type
3:05	30,31,33	30,31,33	both teams created 'mirrors' and are mirroring each other into oblivion
3:22	1,3,4,5 loop	1,3,4,5 loop	groups seeing different things - starting to talk more, confused
3:45		1,3,4,5 loop	hint provides a step forward - into other deceptions
3:50	1,11,12,15		trying DoS against correct network
3:57	1		DoS only partly worked, return to retry
4:00			

Table 4.9: Experiment 2 - Observed Behaviors

Predicted	Observed	Comment
0 OUTSIDE	yes	Obvious OUTSIDE
1,2 loop	1,2 loop	0:24, 0:30, 0:36
1,13 loop	1,13 loop	0:24
	1,2,4	0:24 - no data led to doubt results
1,3,4,5 loop	1,3,4,5	0:49
1,3,4,6	1,3,4,6	1:03
6,7 loop	no	never got to it
6,8 loop	no	never got to it
8,9	no	never got to it
*,1	yes	all the time
0 DMZ	yes	Obvious DMZ 1:15, 1:27
1,2 loop	1,2 loop	2:14
1,13 loop	no	never got anywhere anything real
1,3,4,5 loop	1,3,4,5	1:20, 1:30, 1:34, 1:39, 1:49, 2:47
1,3,4,6	1,3,4,6	1:20
	30.31.33	1:42 arbitrary action with no direction or effect - *blowing off steam*
6,7 loop	no	never got to it
6,8 loop	no	never got to it
7,8 loop	no	never got to it
8,9	no	never got to it
9.10	no	never got to it
8,10	no	never got to it
*,1	yes	all the time
0 INSIDE		Obvious INSIDE 2:00, 2:47
1,2 loop	no	never got to it
1,13 loop	no	never got to it
1,3,4,5 loop	1,3,4,5 loop	2:56, 3:22, 3:45
1,3,4,6	1,3,4,6	2:51
6,7 loop	no	never got to it
6,8 loop	6,8,5,1	2:51
7,8 loop	no	never got to it
8,9	no	never got to it
9.10	no	never got to it
8,10	no	never got to it
1,11,12,18		never got to it
1,11,12,15	1,11,12,15	2:29
15,16 loop	no	never got to it
15,17	no	never got to it
16,17	no	never got to it
15,20	no	never got to it
30,34,35 loop	no	never got to it
30,31,33 loop	30,31,33	3:05
*,1	yes	all the time

Table 4.10: Experiment 2 - Results

Time	Sequence	Comment
0		Configure network - start in Outside network
0:10	1,3,4,5 loop	Thought they found computers but were confused
	1,3,4,6,7 loop	Occasionally thought something was real but then not
	1,3,4 loop	Thought equipment might be bad.
	1,2 loop	Never found many real targets
	1,13 loop	Never found several false targets
5:30	==> DMZ	All decide to move to DMZ network
	1,2 loop	Never found real targets
	1,13 loop	Never found several false targets
	1,3,4,5 loop	
	1,13,1 loop	
	1,11,12,18 loop	
	1,11,12,15,16,18 loop	
	1,11,12,15,16,8 loop	
	30,31,33	Frustration led to random attempts at exploits
9:00		END OF TIME

Table 4.11: Experiment 3 - Observed Behaviors

4.8 Experiment 3

In experiment 3, we repeated experiment 2 under somewhat different conditions. In this case, 9 hours were provided for the attackers. Attack groups included volunteers at a conference who were attending classes in attacking computer systems and participants in a contest wherein they could win thousands of dollars in prizes for defeating the defenses. We used the same predictions for this experiment as for experiment 2.

The behaviors shown in Table 4.11 were observed over a 9 hour period of attempted entry. Times were not accurately kept because the situation was less amenable to close observation. The results are summarized in Table 4.12.

In this experiment, it seems clear that the attackers were less able to make progress. This appears to have a great deal to do with the level of experience of the attackers against the defenses in place. Despite having more than twice the available time, the attackers were unable to penetrate many of the deceptions at all, and were unable to succeed even against simple targets. It took nearly 8.5 hours before attackers started taking detailed notes of all the things they saw in order to try to correlate their observations. By comparison, students in previous experiments who had been trained in red teaming against deceptions in earlier efforts started taking notes immediately.

The only unpredicted behavior was the movement toward attempts at random exploits (i.e., 30.31.33). It appears that this results from frustration in other areas. This is particularly important because we had anticipated that such things could happen but did not understand the circumstances under which it might happen. We now believe that we have a better basis for understanding this and that we will be able to specifically generate conditions that induce or prevent this behavior.

4.9 Summary, Conclusions, and Further Work

It appears, based on this limited set of experiments, that in cases wherein attackers are guided by specific goals, the methods we identified in this and previous papers can be used to intentionally guide those attackers through desired paths in an attack graph. Specifically, the combination of

Predicted	Observed	Comment
0 OUTSIDE	yes	Obvious OUTSIDE
1,2 loop	1,2 loop	
1,13 loop	1,13 loop	
1,3,4,5 loop	1,3,4,5 loop	
1,3,4,6	1,3,4,6	
6,7 loop	6,7 loop	
6,8 loop	no	never got to it
8,9	no	never got to it
*,1	yes	all the time
0 DMZ	yes	Obvious DMZ
1,2 loop	1,2 loop	most of the time
1,13 loop	1,13 loop	several of them
1,3,4,5 loop	1,3,4,5 loop	Much of the time
1,3,4,6	no	never got to it
	30.31.33	Frustration led to random attempts at exploits
6,7 loop	no	never got to it
6,8 loop	no	never got to it
7,8 loop	no	never got to it
8,9	no	never got to it
9.10	no	never got to it
8,10	no	never got to it
*,1	yes	all the time
	1,11,12,18 loop	Unanticipated, but within the attack graph
	1,11,12,15,16,18 loop	Unanticipated, but within the attack graph
	1,11,12,15,16,8 loop	Unanticipated, but within the attack graph

Table 4.12: Experiment 3 - Results

directed objectives with the induction and suppression of signals that are interpreted by computer and human group cognitive systems leads to the ability to induce specific errors in the group cognitive system leading to guided movement through an attack graph.

The ability to guide groups of human attackers and their tools through deception portions of attack graphs and keep them away from their intended targets appears to provide a new capability for the defense of computer systems and networks. This method appears to operate successfully for periods of 4-9 hours against skilled human attack groups with experience in attack and defense and access to high quality tools and may operate for far longer periods. The number of experiments of this sort is clearly limited to the point where meaningful statistical data cannot be gleaned and further experimental studies are called for to further refine these results.

One area of particular interest is the ability of deceptions of this sort to operate successfully over extended periods of time. It appears that these defenses can operate successfully over time, but it also seems clear that with ongoing effort, eventually an attacker will come across a real system and penetrate it unless these defenses lead to adaptation of the defensive scheme. We foresee a need to generate additional metrics of time and information theoretic results to understand how long such deceptions can realistically be depended upon and to what extent they will remain effective over time in both static and adaptive situations.

Chapter 5

Errors in the Perception of Computer-Related Information

by Fred Cohen and Deanna Koike[1]
Jan 12, 2003

- Fred Cohen: Sandia National Laboratories
- Deanna Koike: Sandia National Laboratories (CCD), UC Davis

5.1 Abstract

This paper describes a set of error types associated with human and machine cognition of sequences of passive observations and observations of responses to stimuli in computer and computer network environments. It forms the foundation of a theory for the limits of deception and counterdeception in attack and defense of computer networks and systems.

5.2 Background and Introduction

As part of the effort to understand the limits of cognition and the issues related to deception and counterdeception in the information arena, this paper examines a model of error types associated with this sort of cognition. The belief in the need for some characterization of error types flows from the progress made in fault tolerant computing when the idea of creating a fault model was first introduced. Initial fault models were based predominantly on stuck-at faults in which memory bits and inputs or outputs of digital logic gates were stuck at either an on 'ON' or an 'OFF' state. While these models were not comprehensive, they did cover a significant portion of the space of real errors in digital systems and they permitted systematic analysis of the fault space which could then be mathematically driven through sets of equations associated with a design in order to create test sets, perform automated diagnosis, and analyze designs and implementations for theoretical error conditions that were found in the real world. In a similar manner, this paper proposes a model of error types associated with the cognitive processes undergone by computer software, humans, and organizations, in the hope that it will lead to useful models and perhaps deeper mathematical understanding of the challenges and solutions associated with deception and counterdeception in this arena.

[1]This chapter is published online at `http://all.net/journal/deception/Errors/Errors.html`.

In Chapter 2 our review of the history of deception and the cognitive issues in deception discussed much of the previous work in understanding issues in human cognition. The wide range of experiments done on perception in the visual, sonic, olfactory, and tactile systems indicates that some set of simple rules can be used to model the cognitive systems underlying these perceptual domains and that specific error types can be induced in these systems by understanding and inducing the misapplication of these rules. For example, the human visual system observes flashes of light associated with photons striking the optic nerve and emits impulses into the brain. These impulses are processed by neural mechanisms to do things like line detection and motion detection. If the line detection mechanism detects two line segments whose end points are coincident in both eyes (or one if the other is disabled), the cognitive system interprets this as a contact between those two lines in 3-dimensional space. Common optical illusions involve the creation of situations in which observers are constrained in their perspective so as to observe things like two line segments whose end points are co-incidental but in which the actual mechanical devices are not touching. An object can then be 'passed through', seemingly by magic.

It is the thesis of this paper that a very similar set of processes exist in the human and machine cognition of information such as network traffic and machine state and that, with sufficient understanding of these cognitive systems, a set of identifiable error types can be selectively and reliably induced by exploiting these processes. Similarly, we hope to find methods to seek out the limits of the complexity of creating deceptions in this manner and of countering deceptions through selective observation and analysis of observables.

Initial experiments on deception indicate that errors can be induced (see Chapter 3 and more recent experimental results indicate that these errors can be induced in such a manner as to drive subjects through specific and predictable behavioral sequences (see Chapter 4. These experiments were undertaken under the assumptions about error types discussed in this paper and thus act as a limited confirmation of the model's validity in terms of its relationship to real-world faults at some level of abstraction.

5.3 A Basic Notion of Observation

We wish to consider attacks by an organized set of actors (people, other creatures, their technologies, and their group interaction mechanisms) against some other organized set of actors. We will call the group forming the system under attack the 'defender' (D) and the group attacking the defender the 'attacker' (A). By this we mean that some set of processes and capabilities are being applied by the attacker to try to understand something (in the semantic sense) about the defender. Interactions between attacker and defender may operate through the rest of the world (W), and actors and their interactions may be part of both the attacker and the defender. We assume that the cognitive systems of attacker and defender and the operation of the rest of the world can be modeled to an adequate degree of resolution by some sort of state machines:

All possible n dimensional state spaces:

$$S^* := S_0^u, S_1^u, S_2^u, ..., \forall S_x^u \in S^* \mapsto S_x^u \in R^n$$

There are an infinite number of possible state spaces, with n-dimensional real ($\aleph_1$- sized) things (perhaps).

$$\begin{array}{lll} S_0^u & := & s_{0.0}^u, s_{0_1}^u, s_{0.2}^u, \dots \\ S_1^u & := & s_{1.0}^u, s_{1.1}^u, s_{1.2}^u, \dots \\ \dots & & \\ \text{where } s_{x.y}^u \in R^n & & \end{array}$$

In this notation, x may be thought of as time and y as the name of the state variable.

We define the Universe as follows:

$$U := (S^u, t, F^u, F_n^u : t \times S_n^u \mapsto S_{n+1}^u, t : F_n^u \times S_n^u \mapsto F_{n+1}^u)$$

The universe can, presumably, change the totality of states and state transitions with time. For convenience, we will drop the 'X_n' notation when a statement applies for all n. We also use '$\rightarrow$' for implication and '$\mapsto$' for a mapping. For the purposes of our discussion, we are concerned with attackers, defenders, and others, each of which comprise subsets of the universe:

The Universe:

$$\begin{aligned} U &: (S_0^a, S_0^d, S_0^w, S_1^a, S_1^d, S_1^w, ...) \in S^* \\ S^u &:= S^w \cup S^d \cup S^a \\ F^u &:= F^w \cup F^d \cup F^a \end{aligned}$$

Interfaces:

$$\begin{aligned} & I^a \in S^a, I^d \in S^d, I^w \in S^w \\ & I^{ad} := I^{da} := I^a \cap I^d \\ & I^{aw} := I^{wa} := I^a \cap I^w \\ & I^{wd} := I^{dw} := I^w \cap I^d \\ & I^a \cup I^d \cup I^w = \emptyset \\ & I^{ad} \cup I^{aw} = \emptyset, I^{wd} \cup I^{wa} = \emptyset, I^{da} \cup I^{dw} = \emptyset \end{aligned}$$

Attacker, Defender, World, and Universe:

$$\begin{aligned} A &:= (\{I^a, S^a, F^a, t\} F_n^a : t \times I_n^a \times S_n^a \mapsto \{S_{n+1}^a, I_{n+1}^a\}, \\ & \quad t : I_n^a \times S_n^a \times F_n^a \mapsto F_{n+1}^a) \\ D &:= (\{I^d, S^d, F^d, t\} F_n^d : t \times I_n^d \times S_n^d \mapsto \{S_{n+1}^d, I_{n+1}^d\}, \\ & \quad t : I_n^d \times S_n^d \times F_n^d \mapsto F_{n+1}^d) \\ W &:= (\{I^w, S^w, F^w, t\} F_n^w : t \times I_n^w \times S_n^w \mapsto \{S_{n+1}^w, I_{n+1}^w\}, \\ & \quad t : I_n^w \times S_n^w \times F_n^w \mapsto F_{n+1}^w) \\ & \forall x \in S_n^u, x \in I_n^d \wedge x \in I_n^a \rightarrow x = x \\ & \forall y \in S_{n+1}^u, y \in I_{n+1}^d \wedge y \in I_{n+1}^a \rightarrow y = y \\ & \forall x \in S_n^u, x \in I_n^d \wedge x \in I_n^{da} \rightarrow x \in I_n^a \\ & \text{and similarly for } A, W \end{aligned}$$

We will represent these as:

$$A_n \mapsto A_{n+1}, D_n \mapsto D_{n+1}, W_n \mapsto W_{n+1}$$

A sequence of k steps is represented with k over the arrow:

$$A_n \overset{k}{\mapsto} A_{n+k}, D_n \overset{k}{\mapsto} D_{n+k}, W_n \overset{k}{\mapsto} W_{n+k}$$

which may also be anotated as F^k.

Also note that I^{ad}, I^{aw}, and I^{wd} directly constrain possible F^a, F^d, F^w.

The 'Interface' states (I^w, I^a, and I^d) are shared states between state spaces. Thus communication consists of changes in the shared states. Talking is different than touching. In touching communication is direct via the shared portion of space at the interface between the parties. In talking, communication is indirect in that the shared portion of state that changes at the inteface between the talker and the world results in state changes in the world which result in state changes at the shared states between the world and the listener.

The actors and their interaction technologies are physical sets of objects in a multidimensional (we normally think in terms of three of them), unlimited and ever-changing universe with infinite granularity. Thus everything is of size at least $\aleph - 1$ in three spatial dimensions and 't' (time or more generally some 'thing') changes the elements of those spaces as well as their states. While a more elaborate model of the universe (perhaps using super string theory) might be technically more accurate, this will do for our purposes. The topology, membership, and state of the elements of the subspace of the universe comprising the attacker, defender, and world also change over time. The attacker, defender, and world may share some parts of the universe, which we will call their interfaces (I).

The human visual system observes flashes of light that strike the eyeball (i.e., state changes at the interface between the world and the actor). It uses cognitive mechanisms to turn those flashes into sets of semantic entities such as representations of chairs and tables within the context of visualization. Similarly, the attacker uses its cognitive mechanisms to turn observables (i.e., state changes at the interface) into sets of semantic entities such as types of information systems and communications protocols within the context of its attacks. In terms of the state machine description: $I^{aw}_{n-1} \mapsto I^{aw}_n \wedge S^a_n \overset{k}{\mapsto} S^a_{n+k}$. But in terms of what we are discussing, we need the additional notion of models.

The 'static case' with respect to U, A, D, and W assumes that U does not change with t (i.e., $S^u_n = S^u_{n+1}$ and $F^u_n = F^u_{n+1}$) and that $A \in U$, $D \in U$, and $W \in U$ do not change their state spaces (i.e., $A_n = A_{n+1}, D_n = D_{n+1}, W_n = W_{n+1}$). In this case, A, D, and W can be treated as state machines wherein the total set of states and state transitions for each remains the same and only their state values change. Other static cases may arise, but this one is particularly useful because it represents the most easily analyzed case.

Static case w.r.t. U, A, D, and W:

$$\begin{aligned}
A &:= (\{I^a, S^a, F^a\}, F^a : I^a \times S^a \mapsto \{S^a, I^a\}) \\
I^a &:= (S^a \cap (S^d \cup S^w)) \\
D &:= (\{I^d, S^d, F^d\}, F^d : I^d \times S^d \mapsto \{S^d, I^d\}) \\
I^d &:= (S^d \cap (S^a \cup S^w)) \\
W &:= (\{I^w, S^w, F^w\}, F^w : I^w \times S^w \mapsto \{S^{w'}, I^{w'}\}) \\
I^w &:= (S^w \cap (S^a \cup S^d))
\end{aligned}$$

Within the general framework, we can describe the basics of observation. We start with direct observation. In direct observation, there are shared states between A and D and the values and changes in values of those states are directly available to A and D. In the case of an attacker observing a defender, the process goes like this:

$$s^d_{n.m} \in I^{da}_n = s^a_{n.m} \in I^{da}_n$$

To the extent that state changes in D are reflected in state changes in I^{da} and that those changes are reflected in I^a, we can add additional steps to the process as follows:

$$S^d_x \overset{y}{\mapsto} I^{da}_{x+y} \overset{z}{\mapsto} I^a_{x+y+z}$$

In the case of indirect communication, the world intervenes and the process expands to:

$$S^d_x \overset{y}{\mapsto} I^{dw}_{x+y} \overset{z}{\mapsto} I^w_{x+y+z} \overset{a}{\mapsto} S^w_{x+y+z+a} \overset{b}{\mapsto} I^{wa}_{x+y+z+a+b} \overset{c}{\mapsto} S^a_{x+y+z+a+b+c}$$

5.4 Models and Model Errors

In general, for each of the 'items' U, A, D, W, $a \in A$, $d \in D$, and $w \in W$, cognitive systems are mechanmisms for making mappings between real situations and models of situations. Each of these items can stay constant or change as a function of t and F. We notate constancy '$[\cdot]$', change '$\langle\cdot\rangle$', realities '$[\cdot]_r$, $\langle\cdot\rangle_r$', models '$[\cdot]_m$, $\langle\cdot\rangle_m$', mappings '$\sim$', exact matches '='.

A model is a mapping from the set of 'real' states, state spaces, functions, and times of interest into a set of 'model' states, state spaces, functions, and times. Models are created for some purpose. The suitability of the model to the purpose and the accuracy with which the 'desired' mapping is

Symbols	Meaning
$[U]_r$	The state of the real universe
$\langle U\rangle_r$	Changes in the real universe
$[U]_m$	The state of a model of the universe
$[U]_r \sim [W]$	A mapping from $[U]_r$ to $[U]_m$
$[A_n \mapsto A_{n+1}]_r$	The real constancies of A over t from n to $n+1$
$\langle A_n \mapsto A_{n+1}\rangle_m$	A model of changes of A over t from n to $n+1$
$\langle F_n \mapsto F_{n+1}\rangle_r \sim \langle F_n \mapsto F_{n+1}\rangle_m$	The mapping from real changes in F from n to $n+1$ into a model of those changes.

Table 5.1: Summary of the Model

met by the 'actual' mapping are functions of the model in use. The symbols and their meanings in the model are summarized in Table 5.1.

Since models in our case are subsets of A and D, the following are always guaranteed to be true:

$$[U]_r \neq [U]_m, \langle U\rangle_r \neq \langle U\rangle_m$$

That is, the reality of the universe is not identical to the model of the universe because the model cannot retain enough states to be precise. As a result, constancies and changes in the real universe do not always get reflected in constancies and changes in the model.

For cases where A and D are smaller than W (certainly the case in most situations), for the same reason as for the case of the whole of U, the following are also true for both A and D:

$$[W]_r \neq [W]_m, \langle W\rangle_r \neq \langle W\rangle_m$$

Similarly, because of the nature of uncertainty about knowledge of states of the universe, changes in the reality cannot be perfectly reflected in changes in the model:

$$\text{for } A : [D]_r \neq [D]_m, \langle D\rangle_r \neq \langle D\rangle_m$$

$$\text{for } D : [A]_r \neq [A]_m, \langle A\rangle_r \neq \langle A\rangle_m$$

The total number of possible models is limited only by the number of sequential machines possible within A or D, the maximum complexity of F, and t. For general purpose mappings, this is the power set of the number of overall state values, but physics limits state transitions to physically proximate states, so this is only an upper bound. Clearly we cannot explore all possible models (as defenders, our exploration of models is part of the general effort to model $A \in D$), but we can look at what we believe to be a fruitful set of models of situations we observe in the world.

With this addition of modeling we can make a more useful characterization of the processes of direct and indirect observation and experimentation and start to talk about the nature of errors in models as well as in the underlying physics and mathematics of systems with states. We begin with the general characterization of errors.

All cognitive systems are limited. Whether it is a result of finite memory, time, granularity, observables, performance, operational range, design, or other factors, errors are possible in all such systems. At identified cognitive 'levels' (as defined in Chapter 2 of current and anticipated systems, we can identify specific errors and error types. A complete set of 'errors' relative to a model can be constructed of differences between the 'desired' and 'actual' mappings between items and the models of those items, both in terms of their states and state spaces and changes in their states and state spaces. We then write the general set of errors in terms of the 'desired' mappings as:

$$\mathit{Errors} := \{i | i \in \{U, A, D, W, a \in A, d \in D, w \in W\} \wedge [i]_m \not\sim [i]_r \wedge \langle i\rangle_m \not\sim \langle i\rangle_r\}$$

In other words, the complete set of possible errors consist of all cases in which the mapping of constancy and change of states, state spaces, functions, and times in the real system into desired equivalent constancy and change of states, state spaces, functions, and times in the model are not desired mappings.

5.5 Passive Observation

Assume that the attacker seeks to covertly collect observables and fuse these observations into a semantic model of what is taking place in the defender without inducing any signals into (or responses by) the defender. This may either be a case of direct observation:

$$I_n^{ad}$$

or indirect observation:

$$I_n^{dw} \overset{k}{\mapsto} S_{n+k}^{w} \overset{j}{\mapsto} I_{n+k+j}^{wa}$$

As a notational convenience we may write this as:

$$I^d \mapsto S^a$$

with the special cases:

$$\langle I^{da} \rangle_r$$

and

$$\langle I^{dw} \rangle_r \mapsto \langle S^w \rangle \mapsto \langle I^{wa} \rangle_r \mapsto \langle S^a \rangle$$

We can characterize the situation using the diagram in Figure 5.1.

Observables $[I_k^a]_r, \langle I_k^a \rangle_r$ are processed as a series of time slices

$$(\{\langle S_k^a \rangle_r, [S_k^a]_r\}, \{\langle S_{k+1}^a \rangle_r, [S_{k+1}^a]_r\}, \ldots)$$

that are formed into a model

$$(\{\langle S_k^a \rangle_m, [S_k^a]_m\}, \{\langle S_{k+1}^a \rangle_m, [S_{k+1}^a]_m\}, \ldots)$$

of activities over time. A typical model for IP traffic is based on sets of 'sessions'

$$\{([s_{i.x}^a]_m, [s_{j.x}^a]_m, [s_{k.x}^a]_m)([s_{i.y}^a]_m, [s_{j.y}^a]_m, [s_{k.y}^a]_m), \ldots\}$$

some of which may be related. This model comes from the syntax and semantics of the IP protocols. While other models might have utility, it seems obvious, even if it is not true, that minimizing the difference between $\{\langle X \rangle_m, [X]_m\}$ and $\{\langle X \rangle_r, [X]_r\}$ would tend to give more precise model of X and that this tends to be minimized if the model is forced to follow the syntax and semantics dictated by the nature of the protocol in use. In mathematical terms, we notion that for any metric of the quality of a model:

$$(\{\langle X \rangle_r, [X]_r\} \not\sim \{\langle X \rangle_m, [X]_m\}) < (\{\langle X \rangle_r, [X]_r\} \sim \{\langle X \rangle_m, [X]_m\})$$

In other words, we assert implicitly in the use of this model that a situation in which there is a mapping from reality to the model is never worse for the modeller than a situation in which there is not a mapping from reality to the model. We cannot, at this point, prove any such thing, however, assuming we are talking about 'desired' mappings, our definition of an error (from above) is:

$$[i]_m \not\sim [i]_r \vee \langle i \rangle_m \not\sim \langle i \rangle_r$$

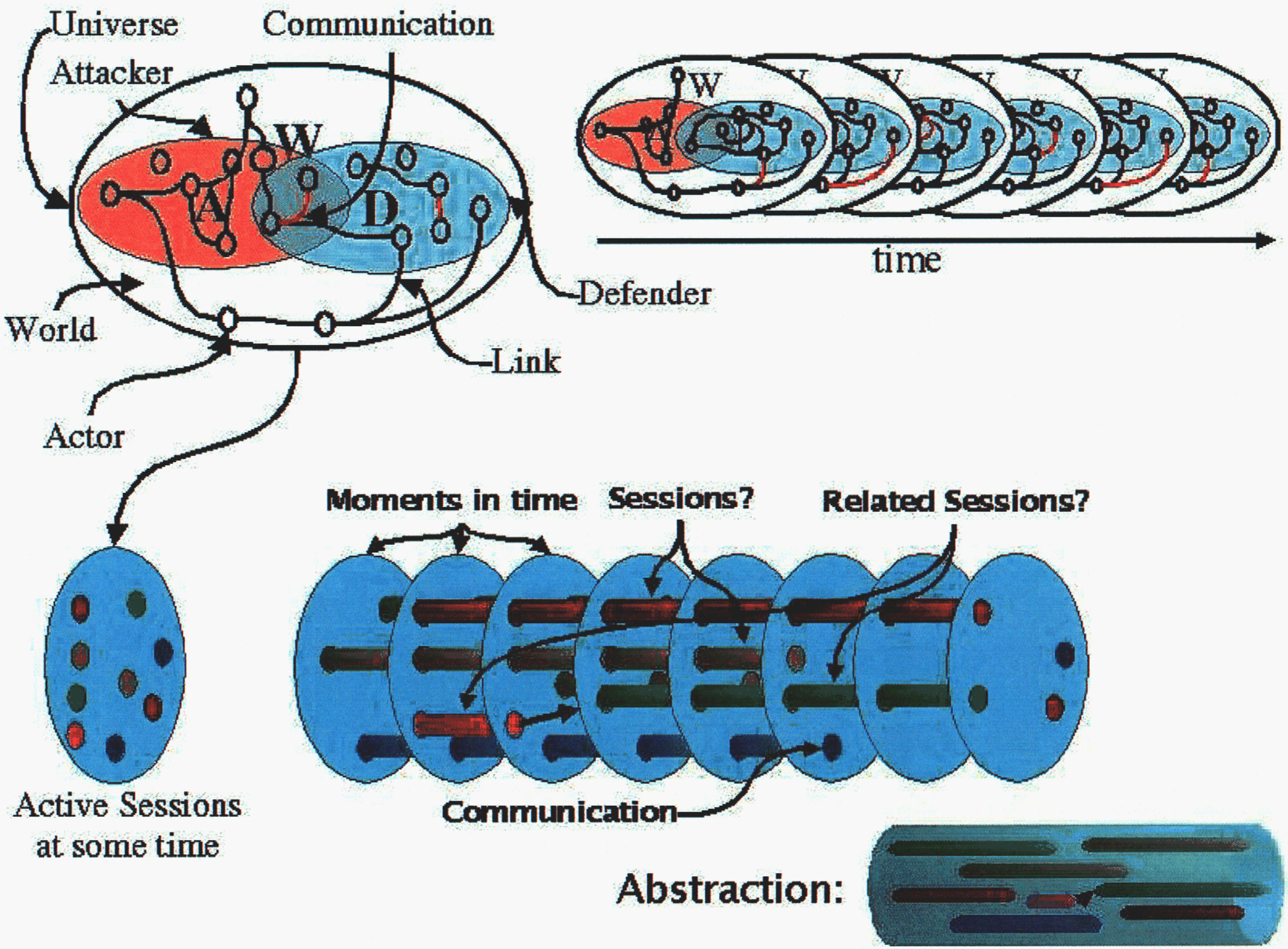


Figure 5.1: Attacker's Passive Observation

thus, there are attacker errors whenever:

$$[S_k^d]_r \not\sim [S_k^a]_m \vee \langle S_k^d \rangle_r \not\sim \langle S_k^a \rangle_m$$

By the nature of the process described above, the time delay associated with transmission of information from D to A guarantees that there is a time lag between some element(s) $\langle S^d \rangle_r$ and $\langle S^a \rangle_m$ unless $D \subset A$ or A is so good at modeling D that the model always correctly anticipates the next observable. But in this case, no observation is necessary because A always accurately preducts changes and state associated with its model of D. There are cases in which real attacks have models that ignore observables from D, however, these rarely result in success in any real sense for attackers.

The processes above provide for several different opportunities for errors. Specifically, for I^{da}:

$$[I^d]_r = [I^a]_r \wedge \langle I^d \rangle_r = \langle I^a \rangle_r$$

Thus the only opportunities for errors are in:

$$[I_k^d]_r \overset{j}{\mapsto} [S_{k+j}^a]_m \wedge \langle I_k^d \rangle_r \overset{j}{\mapsto} \langle S_{k+j}^a \rangle_m$$

Possible error sources include time-related errors (latency), timing-related errors (jitter and misordering), inaccurate mappings, made or missed constancies and changes, and modeling fidelity errors are all possible. Assuming that the model makes sessions, relations, and content, errors also include make and miss sessions, make and miss associations, and make and miss content. All of these can be modeled in terms of constancy and changes in the transforms (F^j) relative to the desired model. For example, $[I_k^d]_r \overset{j}{\mapsto} [S_{k+j}^a]_m$ is the same as the sequence shown in Table 5.2.

Other than the requirements of equality of identical elements of the universe, all mappings ($f \in F$) have the potential to induce errors. In the simplest case, even a function for duplication of a state can fail to produce equality because of uncertainty. Far more complex sorts of errors can occur. Of course the problem with an 'error' equation is that whether or not a transform is an error depends on the context in which it is applied. If the transform done by F is 'desired' by the party depending on it, it is not in error. Otherwise it is. Thus the same transform is an error or not an error depending on the context it is applied to. In the case of our model of passive observation, we have specifics in terms of errors being any mapping into a model that is not reflective of the meaningful states and state changes of interest to the observer.

Sequences of errors in F^* may compound those errors. In addition, even if F^* does not include errors of identity, the part of the transform from reality to models ($\sim$) represent portions of F^* that might not meet desired F^*. Furthermore, assuming that A builds a model of S^d, D can induce specific deceptions into I^d that are not differentiable from the internally meaningful content within D. In other words, as long as $S^d \neq I^d$, D can induce 'deception' states into I^d via $S^d \mapsto I^d$ that are indifferentiable by A from 'meaningful' states that result in $S^d \mapsto I^d$.

For $I^{dw} \mapsto S^w \mapsto I^{wa} \mapsto S^a$ errors are possible through:

$$[I_i^{dw}]_r \overset{j}{\mapsto} [S_{i+j}^w]_r \overset{k}{\mapsto} [I_{i+j+k}^{wa}]_r \overset{l}{\mapsto} [S_{i+j+k+l}^a]_m \wedge \langle I_i^{dw} \rangle_r \overset{j}{\mapsto} \langle S_{i+j}^w \rangle_r \overset{k}{\mapsto} \langle I_{i+j+k}^{wa} \rangle_r \overset{l}{\mapsto} \langle S_{i+j+k+l}^a \rangle_m$$

We still have the same sorts of errors except that these errors are made possible at more steps along the way.

The attacker wishing to gain increasing semantic value from a sequence of observations might combine observables into increasingly complex models of activities. This may also be the source of errors of the same sorts. Placements associated with attacker sensors (i.e., the size of I^d relative to S^d) and capabilities of sensors, communications, and modeling (i.e., and the extent to which F accurately maps predecessor to successor) may limit the accuracy of these processes and models.

A contemporary example of this sort of attack mechanism is the typical network analyzer that takes a series of network packets from one observation point and fuses them together into a series of

$$(I_k^d, S_k^d, I_{k+1}^d, S_{k+1}^d, F_k^d : I_k^d \times S_k^d \times F_k^d \mapsto S_{k+1}^d, I_{k+1}^d, F_{k+1}^d)$$
$$(I_k^a, S_k^a, I_{k+1}^a, S_{k+1}^a, F_k^a : I_k^a \times S_k^a \times F_k^a \mapsto S_{k+1}^a, I_{k+1}^a, F_{k+1}^a)$$
$$(I_k^w, S_k^w, I_{k+1}^w, S_{k+1}^w, F_k^w : I_k^w \times S_k^w \times F_k^w \mapsto S_{k+1}^w, I_{k+1}^w, F_{k+1}^w)$$
$$\forall x \in S_k^u, I_{k.x}^d \in I_{k.x}^{da} \to I_{k.x}^a$$
$$\forall y \in S_{k+1}^u, I_{k+1.y}^d \in I_{k+1.y}^{da} \to I_{k+1.y}^a$$
$$\forall x \in S_k^u, I_{k.x}^d \in I_{k.x}^{dw} \to I_{k.x}^w$$
$$\forall y \in S_{k+1}^u, I_{k+1.y}^d \in I_{k+1.y}^{dw} \to I_{k+1.y}^w$$
$$\forall x \in S_k^u, I_{k.x}^w \in I_{k.x}^{wa} \to I_{k.x}^a$$
$$\forall y \in S_{k+1}^u, I_{k+1.y}^w \in I_{k+1.y}^{wa} \to I_{k+1.y}^a$$

$$(I_{k+1}^d, S_{k+1}^d, I_{k+2}^d, S_{k+2}^d, F_{k+1}^d : I_{k+1}^d \times S_{k+1}^d \times F_{k+1}^d \mapsto S_{k+2}^d, I_{k+2}^d, F_{k+2}^d)$$
$$(I_{k+1}^a, S_{k+1}^a, I_{k+2}^a, S_{k+2}^a, F_{k+1}^a : I_{k+1}^a \times S_{k+1}^a \times F_{k+1}^a \mapsto S_{k+2}^a, I_{k+2}^a, F_{k+2}^a)$$
$$(I_{k+1}^w, S_{k+1}^w, I_{k+2}^w, S_{k+2}^w, F_{k+1}^w : I_{k+1}^w \times S_{k+1}^w \times F_{k+1}^w \mapsto S_{k+2}^w, I_{k+2}^w, F_{k+2}^w)$$
$$\forall x \in S_{k+1}^u, I_{k+1.x}^d \in I_{k+1.x}^{da} \to I_{k+1.x}^a$$
$$\forall y \in S_{k+2}^u, I_{k+2.y}^d \in I_{k+2.y}^{da} \to I_{k+2.y}^a$$
$$\forall x \in S_{k+1}^u, I_{k+1.x}^d \in I_{k+1.x}^{dw} \to I_{k+1.x}^w$$
$$\forall y \in S_{k+2}^u, I_{k+2.y}^d \in I_{k+2.y}^{dw} \to I_{k+2.y}^w$$
$$\forall x \in S_{k+1}^u, I_{k+1.x}^w \in I_{k+1.x}^{wa} \to I_{k+1.x}^a$$
$$\forall y \in S_{k+2}^u, I_{k+2.y}^w \in I_{k+2.y}^{wa} \to I_{k+2.y}^a$$

$$\ldots$$
$$(I_{k+j-1}^d, S_{k+j-1}^d, I_{k+j}^d, S_{k+j}^d, F_{k+j-1}^d : I_{k+j}^d \times S_{k+j-1}^d \times F_{k+j-1}^d \mapsto S_{k+j}^d, I_{k+j}^d, F_{k+j}^d)$$
$$(I_{k+j-1}^a, S_{k+j-1}^a, I_{k+j}^a, S_{k+j}^a, F_{k+j-1}^a : I_{k+j}^a \times S_{k+j-1}^a \times F_{k+j-1}^a \mapsto S_{k+j}^a, I_{k+j}^a, F_{k+j}^a)$$
$$(I_{k+j-1}^w, S_{k+j-1}^w, I_{k+j}^w, S_{k+j}^w, F_{k+j-1}^w : I_{k+j}^w \times S_{k+j-1}^w \times F_{k+j-1}^w \mapsto S_{k+j}^w, I_{k+j}^w, F_{k+j}^w)$$
$$\forall x \in S_{k+j-1}^u, I_{k+j-1.x}^d \in I_{k+j-1.x}^{da} \to I_{k+j-1.x}^a$$
$$\forall y \in S_{k+j}^u, I_{k+j.y}^d \in I_{k+j.y}^{da} \to I_{k+j.y}^a$$
$$\forall x \in S_{k+j-1}^u, I_{k+j-1.x}^d \in I_{k+j-1.x}^{dw} \to I_{k+j-1.x}^w$$
$$\forall y \in S_{k+j}^u, I_{k+j.y}^d \in I_{k+j.y}^{dw} \to I_{k+j.y}^w$$
$$\forall x \in S_{k+j-1}^u, I_{k+j-1.x}^w \in I_{k+j-1.x}^{wa} \to I_{k+j-1.x}^a$$
$$\forall y \in S_{k+j}^u, I_{k+j.y}^w \in I_{k+j.y}^{wa} \to I_{k+j.y}^a$$

Table 5.2: Example Transform Sequence

sessions associated with source and destination IP addresses and TCP, UDP, and other ports and protocols. Some of these analyzers are able to further process this data into IP 'sessions' such as the sequences that are used to fetch electronic mail from servers (e.g., the pop3 protocol). They then provide the means for the human user to request any particular pop3 session and produce a colorized report that shows this specific exchange exclusive of other packets with the communicating parties differentiated by color to facilitate ease of comprehension in the human cognitive system.

Possible failure modes in this process include the limited set of observables associated with the single observation point (i.e., $I^{da} \neq S^{d}$), failures in presentation of observed data to the computer observing the interface (i.e., F fails to achieve identity), failures in presenting that data to the user (i.e., F fails to map actual data into meaningful flashes of light for the user also known as inadequate ($\sim$), and failures of the same sort in the human system observing the computer's output. If we add the analytical components to this analysis, we gain failures in sequence reconstruction, reconstruction when no real sequences exist, failure to notice or resolve ambiguities, identification of ambiguities when there are none, and all of the various errors associated with making and missing content so richly available to humans and the cognitive systems they create.

Current packet analyzers typically go this far, but the typical overall attacker goes much further. The human may fuse sessions together into sets of interrelated protocols, analyze details of packet content for system identification, analyze timings and other data for distance, system characteristics, load levels, and so forth. The attacker might view packet content over time to seek to understand the participants and content associated with communications, try to fuse data from different sensors together to get a more complete picture, or try to fulfill other modeling objectives. It is noteworthy that all current tools make the implicit assumption that they operate in the static case with respect to A, D, W, and U described above. In other words, the tools ignore the possibility of changes in these items.

While these additional steps are not typically automated today, they are potentially automatable, and indications of advancing attack technology strongly support the notion that they will be automated some day. Whether the process is undertaken by a packet analyzer, the human analyst, or an organizational process, from our view, the attacker cognitive system is at work.

The sequence of observables can be characterized as the projection of the flow of cognitive processes in the defender into the attackers's observables. Whatever is to be modeled by the attacker can be thought of as a characterization of the cognitive processes of the defender that induced those observables. These two cognitive systems may not be commensurable in that the attacker's model may involve things that are very different from the things involved in the defender's cognitive system. For example, the defender may be emitting random bit sequences because of a hardware error and the attacker may be modeling these as a new encrypted protocol that must be broken through cryptanalysis. This problem of commensurability comes into play when the attacker or defender try to model the cognitive processes of the other in an effort to 'out think' them. The granularity, depth, accuracy, timeliness, and available resources for making the attacker's model result from the capabilities and intent of the attacker. The efficacy of the model may be affected by the defender's success at controlling the attacker's observables.

5.6 Active Experimentation

Now consider the case where an attacker is willing and able to risk increased exposure by inducing signals into the defender. This may have the affect of providing additional state information or altering states in the defender. Feedback from this process may return to the attacker through their observables. This situation is depicted in the Figure 5.2.

By taking actions and observing the results of those actions, the attacker may affect state changes in and enhance their model of the defender. This process can be thought of as an active search of the defender's state space and an attempt to control the state of the defender also resulting in modification of the attacker's state space. Returning to the visual system, learning takes place

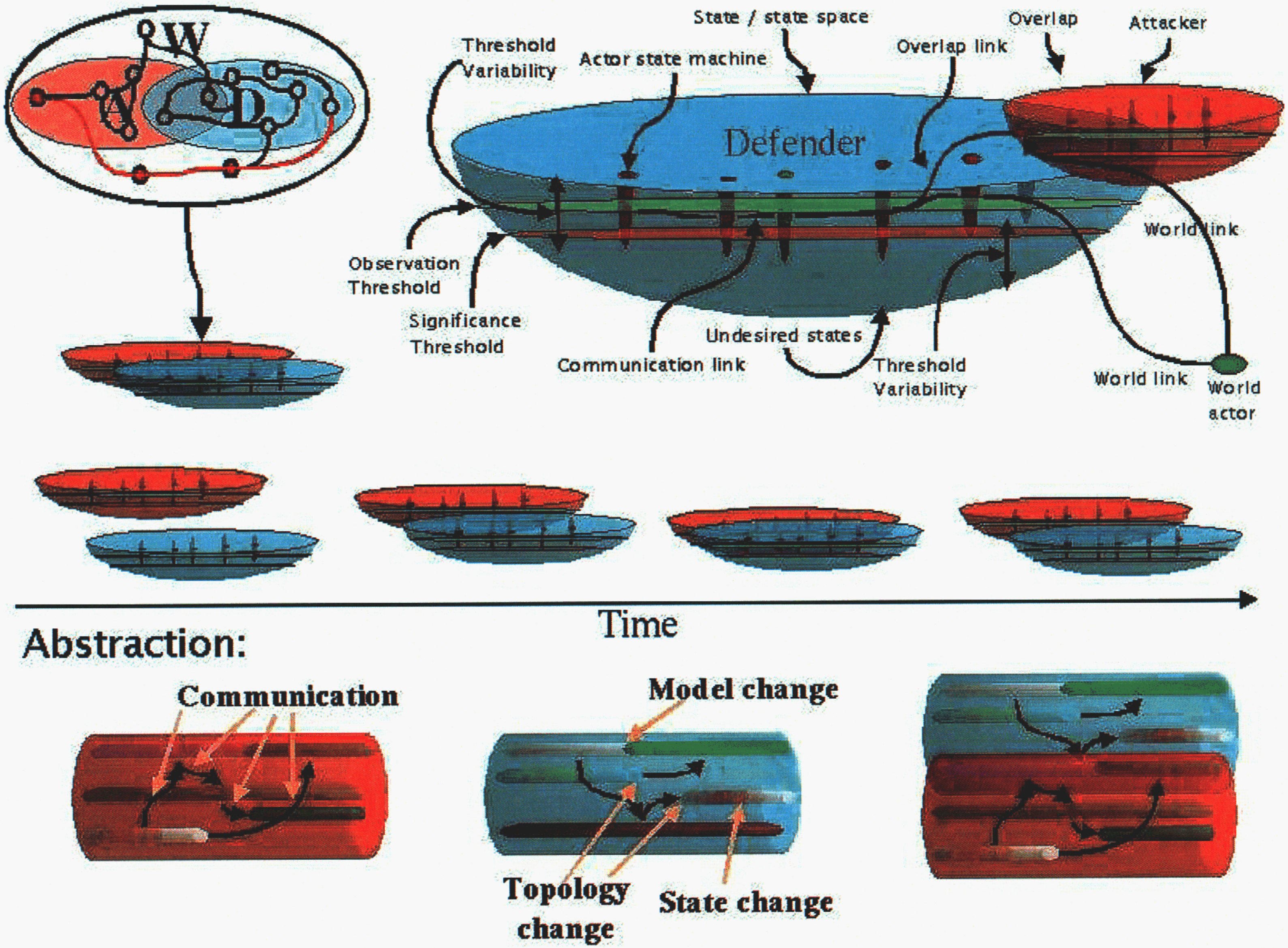


Figure 5.2: Attacker's Active Experimentation

in the brain in the form of rewiring neural connections and changing electrical signal strengths, neural activity levels, frequencies of neuron firings, changes in neurotransmitter levels, and so forth. Similarly, in our model, the attacker, defender and world states and topology may change. The state space and state of the universe may change $t \times S^u \mapsto \{S'^u, S'^{u'}\}$, and attacker, defender, and world subspaces may change $(S^a \times S^d \times S^w \times t) \mapsto \{S'^a, S'^d, S'^w\}$. Relativistic effects will be ignored for now and it will be assumed that the propagation of changes in the universe operate at a rate faster than is of interest to the cognitive systems under consideration.

As in the pure observational case, modeling limitations apply. The total state space of the defender cannot be explored and retained by the attacker unless the attacker has enough available state storage in addition to its own state machine operational needs to store all states and transitions associated with its model of the defender. The model may not be reflective of the true nature of the defender, or may not be adequate to fulfil the objectives of the attacker. The resolution cannot be adequate to create a perfect model except under very limited and unrtealistic ciccumstances. All of the issues that apply to the observation only case apply to the active case as well.

In searching the space, the attacker induces signals that affect the defender. If these are among the set of signals that can be observed by the defender's cognitive system in its state at the time of signal arrival, we call them observables. As the defender's state space is more thoroughly explored by the attacker, some signals and state sequences may be recognized by the defender's cognitive system as requiring specific action. This may trigger alterations in the defender's state, the reachable portions of the defender's state space, the defender's recognition system, and the observables available to the attacker and the defender. There may also be arbitrary delays between the occurrence of an observable and cognitively triggered actions associated with it.

The attacker and defender may also have overlapping elements and those elements may change over time. For example, communications may stop or start on different links, that attacker may take partial control of a defender's computer, there may be insiders planted in the defender's organization, elicitation, deflection, and so forth.

The topology of the attacker, defender, and world may change. New computers might be bought, deployed, positions changed, systems removed, employees hired, fired, changed, children born, death and failure of people and parts, and so forth.

These notions are different from the typical notions of state machine theory in that we are associating semantic differences between portions of the state space. This may be modeled as the division of the state space into subspaces in which different transition sets are applied, but strictly speaking, the division of the state space in such a manner is only useful to the extent that it grants us convenience for our characterization and analysis of the state machines.

As a reminder, the attacker and defender we are speaking of may be the combination of humans, other living creatures, and technologies, so the notion of semantic differences is sensible. Notions like changes in observables clearly apply to these systems because, as an example, the physiology of humans is such that detection thresholds for observables change based on changes in chemical compositions at sensor sites and neurotransmitters at neurons. Furthermore, the mechanisms we are discussing are not necessarily finite state machines. They involve continuous functions in complex and imperfect feedback systems. The brain literally rewires itself, and we are not anthropomorphizing when we bring up things like intent.

In the same manner as the attacker may affect the defender, the defender may affect the attacker through its actions. Presumably, it is the intent of the defender's cognitive system to optimize its overall function. To the extent that this is at odds with the attacker's objectives, these alterations to defender state may have detrimental affects on the attacker's ability to carry out its objectives. It is also possible that the defender's objectives are not at odds with the attacker's and that these state alterations will work to the attacker's benefit.

The result is, potentially, a battle between two cognitive systems; that of the attacker and that of the defender. The nominal objective of the attacker is to gain an adequately accurate model of the defender to induce desired states. The objective of the defender is to prevent undesired state changes in itself and the attacks, and assuming that vulnerabilities exist in its state machines and that the attacker may eventually encounter and exploit them, to affect the cognitive system of the attacker so as to control the cost of such success. We may refer to the purely observational process as a 'passive' attack and the process involving the induction of signals by the attacker into the defender as an 'active' attack.

5.7 Summary of Errors in Cognition

All cognitive systems are limited. Whether it is a result of finite memory, time, granularity, observables, performance, operational range, design, or other factors, errors are possible in all such systems. At the cognition levels of current and anticipated systems, we can identify specific errors and error types, regardless of the mechanisms involved (see Figure 5.3). The complete set of 'errors' can therefore be considered to consist of differences between the desired and actual mappings between items and the models of those items, both in terms of their states and state spaces and changes in their states and state spaces. We then write the general set of errors as:

$$Errors := \{i | i \in \{U, A, D, W, a \in A, d \in D, w \in W\} \wedge [i]_m \not\sim [i]_r \wedge \langle i \rangle_m \not\sim \langle i \rangle_r\}$$

We start with the passive observation case for IP traffic and similar phenomena.

Observables can fail to reflect the total and actual content. In general this can be considered as a composition of missing existing data and making nonexistent data. In all cases of interest some data is missed, if only because of the inability to place unlimited granularity sensors at all points in the physical space of the defender. Thus all of the cognitive process can be seen in a similar light to black box testing. The lack of perfect observables leads to the unavoidable use of assumptions and expectations which in turn results in 'made' data.

The cognitive systems must make assumptions about the nature of the space between sensors in order to build a model. These assumptions combined with the qualitative and quantitative limits of the cognitive system produce the potential for errors in the form of a mismatch between the cognitive model and the real situation. These mismatches can be considered as a recursive combination of missed and falsely created sessions, associations, inconsistencies, and semantic content (a.k.a. understanding). The complexity limits of timely cognition also leads to limits on the accuracy of analysis.

A missed session is a series of interrelated observables that exists in the real system but are not properly modeled. A falsely created (made) session is a model of observables when no such relations exist. A missed association is a relationship between observables in the real system that is not cognitively modeled. A falsely created (made) association is a model of a relationship between observables when no such relationship exists. A missed inconsistency is the failure to detect the relationship between a set of observables that, if properly analyzed, would indicate the presence of an error or imperfect deception. A falsely created (made) inconsistency is a model of the presence of an error or imperfect deception which does not actually exist in the system. A misunderstanding is a semantic interpretation that is not in correspondence with the reality. Semantic interpretation introduces recursion because content and context are dictated by the possibly recursive languages and their syntax and semantics. These same sets of errors can occur at all recursive levels of syntax and semantics as can the misunderstanding errors associated with failure to detect a recursion which exists or falsely modeling a recursion when there is none. All of these error types apply to attacker and defender.

In the case of active experimentation, the errors associated with a passive process are augmented by errors associated with the attempt to search the state space of the system being modeled. These errors can be characterized as the making or missing of models or model changes, topologies and topological changes, communications, and states or state changes. It appears that, based on this overall model, these cases cover the set of all errors that can be made.

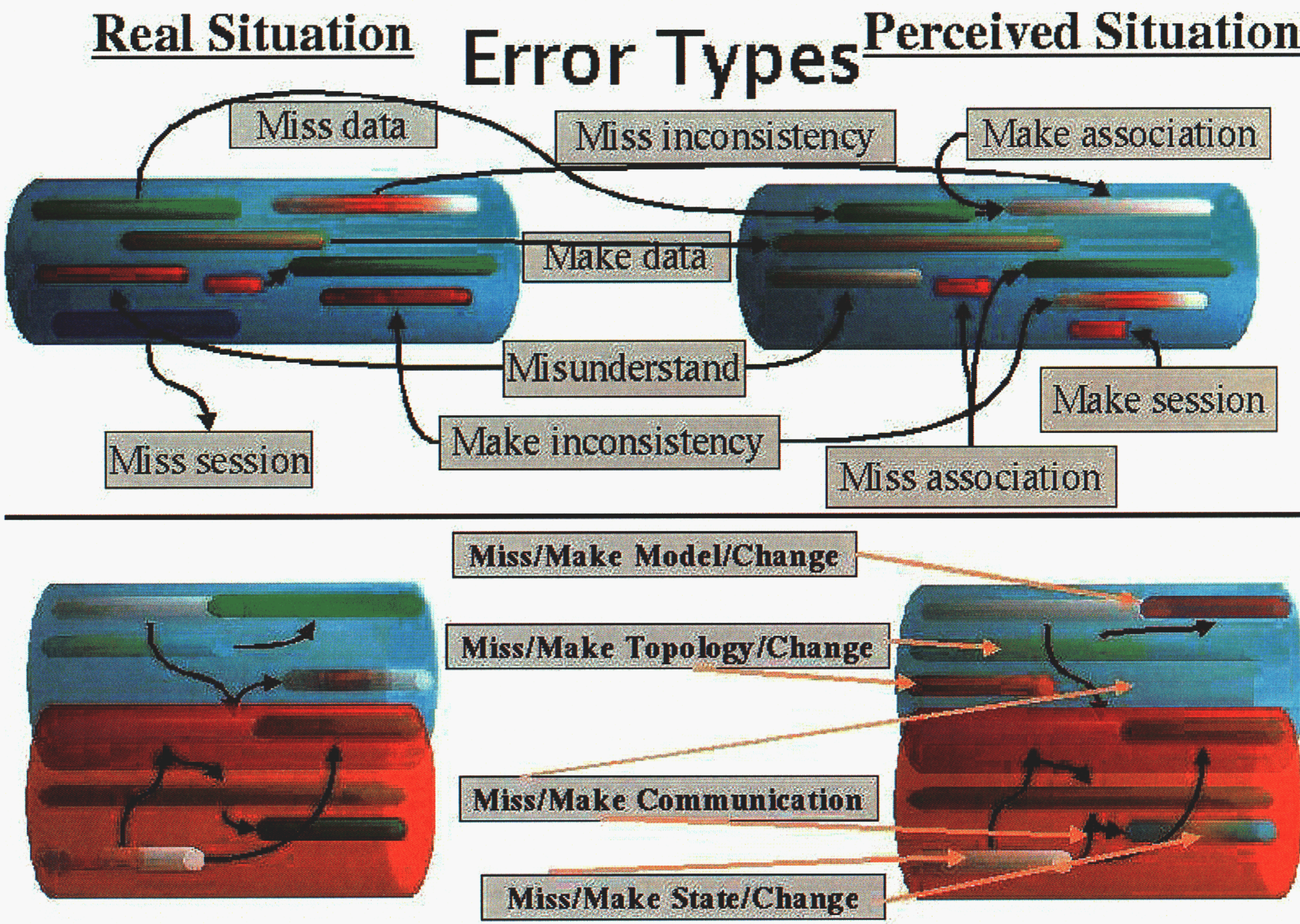


Figure 5.3: Error Types

The cognitive processes associated with exploring a finite machine state space using a black box approach are problematic because of the inherent complexity of black box characterization. It is obvious that it is impossible in general to correctly explore the full state space of a sequential machine from a black box perspective because there may be states that are not reachable once certain input sequences have been encountered. Even if we restrict our interest to machines that can eventually revisit any previous state, the number of possible sequences for such systems is enormous. In the worst case, the length of the sequence required to characterize a finite state machine is exponential in the number of bits of state. Even for large classes of submachines such as those in typical computer languages and processor instruction sets, the size of this set is beyond all hope of exploring in detail. In the case of the systems under discussion, analog as well finite state machines are involved, so appeals to continuity of analog functions are necessary to even approach exploration of these state spaces. Because of digital / analog interactions, aliasing and similar phenomena come into play as well.

Similarly, the enormous number of possible states in most systems limits the ability to assure that only desired states occur. The design of most current computers is such that they are general purpose in function and transitive in their information flow. This guarantees that they are capable of entering many undesirable states and that it is impossible to accurately differentiate all desired from undesired states definitively and in any available amount of time.

The cognitive systems of attacker and defender are limited to non-definitive methods that use limited observables, modeled characteristics, and cognitive processing power to model opponent systems. The challenges in the conflict between attacker and defender are then; (1) for the attacker to select characteristics that provide the desired information and perform a series of observations and experiments that yield the desired states and observables in the defender and (2) for the defender to remain only in desired states and produce only desired interactions. In general, the defender is not limited in retaining its desired states to purely defensive methods. For example, a defender might maintain desired states by actively corrupting the attacker's cognitive model or state. Similarly, the attacker may have to defend itself in order to be effective against such a defender. Indeed, the situation is, at least potentially, symmetrical.

It is entirely possible that the objectives of the attacker and the objectives of the defender are not at odds with each other. In such a case, both the attacker and the defender may be able to operate freely without the need to consider interactions. In other cases, cognitive conflict will be carried out.

5.8 Summary, Conclusions, and Further Work

We have defined and described a model of errors in perception and conditions under which they occur. These error types appear to be unavoidable in any realistic situation in which organizations of humans and computers use cognitive means to analyze computer-related information. This model is a beginning, but surely not the end of this issue.

In the context of cognitive conflict between organizations of humans and computers, this model appears to provide a means by which we may analyze the limits of deception and counterdeception and devise systems and methods with optimal deception and counterdeception characteristics. But we have not yet developed the mathematical results of this theory to the point where we can make practical use of it for the formation of equations or determination of the limits of specific systems or circumstances. Nevertheless, it is a beginning with potential toward this end.

In Chapter 4 some examples were given of specific deception mechanisms that induce specific cognitive errors. In the future, it is anticipated that a far richer set of mechanisms will be available with specific association to cognitive errors they induce, times associated with those errors, and reliability figures related to observation and computation characteristics of the cognitive systems attempting to induce and counter the induction of those errors.

Chapter 6

Conclusions

6.1 Collaboration with UC Davis

One of the goals of this project was to develop a collaborative relationship with the University of California at Davis.

The LDRD funds provided by Sandia to the University of California at Davis were used by the Computer Science department. The funds were used to support students in the deception project, many of whom were advised by Professor Matt Bishop.

In addition to the involvement of UC Davis students in the results outlined in this report, some of the other projects the students worked on were:

- Vicentiu Neagoe, a Masters (now PhD) student, worked on a deceptive Linux kernel (see Trojan Project in Section 6.2.2). He modified several of the system calls so that, when the user is flipped into the deceptive side of the system, the environment is not the true environment, and the user's actions cannot harm the system. He looked at what was necessary to make the deception complete, so the attacker would see the deceptive system as a consistent one (to hide the deception).

- Derek Cotton and Chris Kolina, two UC Davis undergrads, also looked at other aspects of deception, in particular the basic ideas of how to fool the attacker. They also examined basic aspects of the mathematics of deception, including some elementary work on deception as a Turing Machine model, and exchanged ideas with the Sandia group on those lines.

- Deanna Koike Rogers, a Masters student, was a co-author of the papers in Chapter 4 and Chapter 5 and is also involved in the Trojan Project (Section 6.2.2).

6.2 Related Projects

There are several other projects on deception, as well as two patent filings and two provisional patent filings, that are directly related to the deception research described in this report.

6.2.1 Invisible Router (IR)

The Invisible Router (IR) is a network deception tool. As described in the Invisible Router Fact Sheet (SAND2002-8172P [SNL02]), based on a user-programmable set of rules, the IR routes network traffic coming from unauthorized sources to a deception network, while traffic from authorized sources is routed to the production network. The deception network is set up as a mirror of the production network to which the attacker is trying to break in.

The core technology of the IR was used as part of the deception mechanism in the Red-Teaming experiments described in Chapter 3, as well as for other projects for an outside customer.

The IR project has resulted in the filing of patents entitled METHOD AND APPARATUS FOR INVISIBLE NETWORK RESPONDER, METHOD AND APPARATUS FOR SPECIFYING COMMUNICATION INDICATION MATCHING AND/OR RESPONSES and METHOD AND APPARATUS FOR CONFIGURABLE COMMUNICATION NETWORK DEFENSES.

6.2.2 Trojan Project

The Trojan Project is a host-based deception system designed to mitigate the insider threat; that is, it is aimed at an attacker who has penetrated a system and is masquerading as a legitimate user, or at a legitimate user who is attempting to abuse his privileges on a system. Trojan is an ongoing project that includes intrusion detection, decision-making, and deceptive countermeasure components. The focus to date has been mainly on the deception technologies that can be implemented in a Linux-based system. These include an execution wrapper that runs in user-space (with minor modifications to the kernel) to intercept and possibly modify execution calls, and a loadable kernel module that is able to intercept any system call and modify its behavior.

A provisional patent based on the host-based deception technology used in the Trojan project was filed as application 60/416,285 filed 3 Oct 2002 entitled METHOD AND APPARATUS PROVIDING DECEPTIONS AND/OR ALTERED OPERATIONS IN INFORMATION SYSTEMS, with plans to file for two patents by 3 Oct 2003.

Work on the Trojan Project is ongoing, and a SAND report detailing the current state of the project and its future design is under development.

6.2.3 Adaptive Network Countermeasures (ANC)

Adaptive Network Countermeasures (ANC) is a 2-year LDRD project (Project 38744 for FY02 and FY03) that involves the use of deceptions at the network level. This project uses deceptive countermeasures that watch unused subnets and reply to attackers by answering for all addresses and ports as well as by simulating protocol stacks and applications. The project studies techniques to identify attackers, and then mislead and confuse them. See the ANC LDRD SAND Report for more details.

6.3 Conclusions and Future Work

The main goal of this project was to lay the theoretical groundwork for future more applied work in deception. An additional goal was to develop a collaborative relationship with the University of California at Davis. Work in both of these areas is still ongoing, and both the collaborations with UC Davis and the work in deception will continue.

The results of the research presented in this paper seem to support the common notions about deception that are outlined in Section 1.1. The Red Teaming results in Chapter 3 demonstrate that deception techniques have the ability to increase attacker workload and reduce attacker effectiveness. In addition, deception can decrease the defender effort required for detection and provide substantial increases in defender understanding of attacker capabilities and intent. Anecdotal evidence seems to indicate that even belief that deception technology is in use on a system can result in some benefit to the defender, because attackers don't trust their results and do more cross-checking, resulting in slower progress.

The results from Chapter 4 indicate that, at least when attacker goals are known, it is possible to use deception to induce specific errors that guide attackers through a desired attack path.

Chapter 5 describes a model of errors in perception and conditions under which they occur, which may be useful in providing some insight into designing deception systems and methods with optimal deception and counterdeception characteristics, though more work is required to develop the mathematical results of the theory to the point where it can be put to practical use.

Future projects regarding deception will probably be more applied in nature. A new umbrella project called "Anomaly-Response Capability" (ARC), which involves integrated configurable components to 1) identify anomalies, 2) elevate threat levels, and 3) decide, initiate and correlate responses, will include deception technologies as a key component. ARC includes continuations of the Trojan Project (Section 6.2.2) and the ANC Project (Section 6.2.3), as well as other cyber security projects.

The relationship with UC Davis continues, in the form of a new LDRD proposal for FY04 on "Property Based Testing", as well as continuing relationships with UC Davis students at the CCD.

Bibliography

[Arm98] US Army. *Field Manual 90-02: Battlefield Deception.* US Government, 1998.

[Cha01] Daniel Chandler. Personal Home Pages and the Construction of Identities on the Web, 2001. http://www.aber.ac.uk/media/Documents/short/webident.html.

[Che91] Bill Cheswick. An Evening with Berferd, 1991. http://all.net/books/berferd/berferd.html.

[Cia01] Robert B. Cialdini. *Influence: Science and Practice.* Allyn and Bacon, Boston, 2001.

[Coh92] Fred Cohen. Operating System Protection Through Program Evolution. *Computers and Security*, 1992. http://all.net/books/IP/evolve.html.

[Coh96a] Fred Cohen. A Note On Distributed Coordinated Attacks. *Computers and Security*, 1996. http://all.net/books/dca/index.html.

[Coh96b] Fred Cohen. Internet Holes - Internet Lightning Rods. *Network Security Magazine*, July 1996. http://all.net/journal/netsec/1996-07-2.html.

[Coh98a] Fred Cohen. Deception Toolkit, March 1998. http://all.net/dtk/index.html.

[Coh98b] Fred Cohen. Red Teaming and Other Agressive Auditing Techniques. *Managing Network Security*, March 1998. http://all.net/journal/netsec/1998-03.html.

[Coh98c] Fred Cohen. The Unpredictability Defense. *Managing Network Security*, April 1998. http://all.net/journal/netsec/1998-04.html.

[Coh99a] Fred Cohen. A Mathematical Structure of Simple Defensive Network Deceptions, 1999. http://all.net/journal/deception/mathdeception/mathdeception.html.

[Coh99b] Fred Cohen. A Note on the Role of Deception in Information Protection, May 1999. http://all.net/journal/deception/deception.html.

[Coh99c] Fred Cohen. Simulating Cyber Attacks, Defenses, and Consequences, May 1999. http://all.net/journal/ntb/simulate/simulate.html.

[Coh00a] Fred Cohen. Method and Aparatus for Network Deception/Emulation, October 2000. International Patent Application No PCT/US00/31295, Filed October 26, 2000.

[Coh00b] Fred Cohen. The Structure of Intrusion and Intrusion Detection, May 2000. http://all.net/ (InfoSecBaselineStudies).

[Coh00c] Fred Cohen. Understanding Viruses Bio-logically. *Network Security Magazine*, August 2000. http://all.net/journal/netsec/2000-08.html.

[CPS+99] Fred Cohen, Cynthia Phillips, Laura Painton Swiler, Timothy Gaylor, Patricia Leary, Fran Rupley, Richard Isler, and Eli Dart. A Preliminary Classification Scheme for Information System Threats, Attacks, and Defenses; A Cause and Effect Model; and Some Analysis Based on That Model. *The Encyclopedia of Computer Science and Technology*, 1999. http://all.net/journal/ntb/cause-and-effect.html.

[CR03] Fred C. Cohen and Deanna Koike Rogers. Leading Attackers Through Attack Graphs with Deceptions. *Computers and Security*, 22(5), July 2003.

[Deu95] Diana Deutsch. *Musical Illusions and Paradoxes.* Philomel, La Jolla, CA, 1995.

[Dew89] Colonel Michael Dewar. *The Art of Deception in Warfare*. David and Charles Military Books, 1989.

[DH82] Donald Danial and Katherine Herbig, editors. *Strategic Military Deception*. Pergamon Books, 1982.

[DN95] James F. Dunnigan and Albert A. Nofi. *Victory and Deceipt: Dirty Tricks at War*. William Morrow and Co., New York, 1995.

[Far98] Fay Faron. *Rip-Off: a writer's guide to crimes of deception*. Writers Digest Books, Cinn, OH, 1998.

[Fel00] Bob Fellows. *Easily Fooled*. Mind Matters, PO Box 16557, Minneapolis, MN 55416, 2000.

[Gil91] Thomas Gilovich. *How We Know What Isn't So: The fallibility of human reason in everyday life*. Free Press, NY, 1991.

[GRA99] Scott Gerwehr, Jeff Rothenberg, and Robert H. Anderson. An Arsenal of Deceptions for INFOSEC (OUO). Memorandum PM-1167-NSA, National Defense Research Institute Project, October 1999.

[Gre98] Robert Greene. *The 48 Laws of Power*. Penguin Books, New York, 1998.

[Gri78] William L. Griego. Deception - A 'Systematic Analytic' Approach. (slides from 1978, 1983), 1978.

[GWM+00] Scott Gerwehr, Robert Weissler, Jamison Jo Medby, Robert H. Anderson, and Jeff Rothenberg. Employing Deception in Information Systems to Thwart Adversary Reconnaissance-Phase Activities (OUO). Technical Report PM-1124-NSA, RAND National Defense Research Institute, November 2000.

[Han93] Charles Handy. *Understanding Organizations*. Oxford University Press, New York, 1993.

[Heu99] Richards J. Heuer. *Psychology of Intelligence Analysis*. History Staff Center for the Study of Intelligence, Central Intelligence Agency, 1999.

[Hof98] Donald D. Hoffman. *Visual Intelligence: How We Create What We See*. Norton, New York, 1998.

[Hon] The HoneyNet Project. `http://www.honeynet.org`.

[Hub83] Robert E. Huber. Information Warfare: Opportunity Born of Necessity. *Systems Technology (Sperry Univac)*, IX(5):14–21, September 1983.

[HW99] Colonel John Hughes-Wilson. *Military Intelligence Blunders*. Carol & Graf, New York, 1999.

[Ito93] Mimi Ito. Cybernetic Fantasies: Extended Selfhood in a Virtual Community, 1993. `http://www.usyd.edu.au/su/social/papers/ito1.html`.

[Kah67] David Kahn. *The Code Breakers*. Macmillan Press, New York, 1967.

[Kal94] Pamela J. Kalbfleisch. The language of detecting deceit. *Journal of Language & Social Psychology*, 13(4):469–497, 1994. [Provides information on the study of language strategies that are used to detect deceptive communication in interpersonal interactions. Classification of the typology; Strategies and implementation tactics; Discussions on deception detection techniques; Conclusion].

[Kar70] Chester R. Karrass. *The Negotiating Game*. Thomas A. Crowell, New York, 1970.

[Kee93] John Keegan. *A History of Warfare*. Vintage Books, New York, 1993.

[Kno87] C3CM Planning Analyzer: Functional Description (Draft) First Update. Technical Report RADC/COAD Contract F30602-87-C-0103, Knowledge Systems Corporation, December 1987.

[Lam87] D.R. Lambert. A Cognitive Model For Exposition of Human Deception and Counterdeception. Technical Report 1076, NOSC, October 1987. `http://citeseer.nj.nec.com/lambert87cognitive.html`.

[LLN96] Intrusion Detection and Response. National Technical Baseline, Lawrence Livermore National Laboratory and Sandia National Laboratories, December 1996. `http://all.net/journal/ntb/ids.html`.

[Mac89] Charles Mackay. *Extraordinary Popular Delusions and the Madness of Crowds.* Templeton Publications, 1989. (originally Richard Bently Publishers, London, 1841).

[MKU] MKULTRA. A list of documents related to MKULTRA can be found over the Internet.

[MT86] Robert W. Mitchell and Nicholas S. Thompson. *DECEPTION: Perspectives on human and nonhuman deceipt.* SUNY Press, New York, 1986.

[NRC98] *Modeling Human and Organizational Behavior.* National Academy Press, Washington, DC, 1998. National Research Council.

[Pea00] Mark Peace. Dissertation: A Chatroom Ethnography, May 2000. `http://www.aber.ac.uk/media/Students/mbp9702.doc`.

[Pek90] Bob Pekarske. Restoration in a Flash—Using DS3 Cross-connects. *Telephony*, September 1990. [This paper describes the techniques used to compensate for network failures in certain telephone switching systems in a matter of a millisecond. The paper points out that without this rapid response, the failed node would cause other nodes to fail, causing a domino effect on the entire national communications networks].

[RP90] Richard J. Robertson and William T. Powers, editors. *Introduction to Modern Psychology, The Control-Theory View.* The Control Systems Group, Inc., Gravel Switch, Kentucky, 1990.

[Sec00] Al Seckel. *The Art of Optical Illusions.* Carlton Books, 2000.

[SNL02] Invisible Router Fact Sheet, 2002. Sandia National Laboratories – SAND Report SAND2002-8172P.

[SSC] SSCSD Tactical DecisionMaking Under Stress. SPAWAR Systems Center.

[Ste93] Gordon Stein. *Encyclopedia of Hoaxes.* Gale Research, Inc, 1993.

[Tzu83] Sun Tzu. *The Art of War.* Dell Publishing, New York, 1983. Translated by James Clavell.

[Van] Heidi Vanderheiden. Gender swapping on the Net? `http://web.aq.org/~tigris/loci-virtualtherapy.html`.

[Vri00] Aldert Vrij. *Detecting Lies and Deceipt.* Wiley, New York, 2000.

[Wei48] Norbert Weiner. *Cybernetics.* 1948.

[Wes81] Charles K. West. *The Social and Psychological Distortion of Information.* Nelson-Hall, Chicago, 1981.

[Wha69] Bart Whaley. *Strategem: Deception and Surprise in War.* MIT Center for International Studies, Cambridge, MA, 1969.

[Whi97] Chuck Whitlock. *Scam School.* MacMillan, 1997.

[Wil68] Andrew Wilson. *The Bomb and The Computer.* Delacorte Press, New York, 1968.

[WSC] Western Systems Coordinating Council WSCC Preliminary System Disturbance Report Aug 10, 1996 - DRAFT. [This report details the August 10, 1996 major system disturbance that separated the Western Systems Coordinating Council system into 4 islands, interrupting service to 7.5 million customers for periods ranging from several minutes to nearly six hours].

Appendix A

Red Team Standard Pre-Briefing

This is the standard Red Teaming Pre-Briefing Form that was used during the deception exercises described in Chapter 3. An online version of this form is available at:

`http://all.net/journal/deception/experiments/pre-brief.html`

A.1 Introduction

Standard Red Teaming Pre-Briefing

To be reviewed by all participants prior to each exercise.

A.2 Operations Security

This red teaming exercise is a contest between teams and part of a study being done on red teaming and defenses. As such the requirements for operations security are as follows:

- **Threats:**
  - Short term threats include the other red teams (your competitors) until all of the red team exercises are completed and the rest of the world until the results of this research are published (in 6 months to a year typically).
  - Long term threats include those who might exploit what we learn about the defenses we test in order to attack them.
- **Vulnerabilities:**
  - You might tell other people or teams what you or your group are doing and inadvertently:
    1. invalidate the research results,
    2. do less well in the competition,
    3. reveal the study to others who will publish (possibly inferior results) first,
    4. through your interactions with others in other groups you might reveal information that will help their team do better,
    5. reveal details of defensive technologies we are testing to others, and
    6. by discussing results of one exercise with other teams before the end of the whole sequence of red teams you may reveal information that will help them in future rounds of the exercise.

  - Attacks might spill over to other networks and cause harm to the users of those networks.
  - Attacks might harm servers supporting the exercises and damage continued red teaming.
  - While using the web we might inadvertently encounter pornographic content.

- **Mitigation:** In order to mitigate the potential consequences of these threats and vulnerabilities we are taking the following precautions:

  1. The red teaming exercises are being done in a reasonably secure facility from a standpoint of the issues at hand. In addition, to physical security, digital diodes are being used between networks to prevent spillage and physical security of system in the exercise is being increased to prevent accidental cross connects and lightly malicious (a.k.a. overly competitive) behavior.
  2. Don't tell anyone else what your team did or found out until the end of the whole sequence of exercises. Don't tell anyone outside of the CCDs that you are doing this until results are published. Don't tell anyone about any defenses you defeat.
  3. Follow the rules of engagement strictly and do only those things you are permitted to do via these rules, but within the rules, do your best.
  4. DO NOT attempt to defeat any technical protections and do not attack the infrastructure that supports the exercise. Specifically, do not attack the diode or cross connect networks. The former will cause you to be unable to get supporting tools for your efforts, while the latter may be hazardous to your career.
  5. DO NOT use any of the green net systems EXCEPT during these exercises. They get cross connected to other networks during off hours so they can be reloaded for the next run.
  6. All reasonable efforts will be made to avoid pornographic sites. If encountered they will be immediately reported to the observer and at the end of the exercise to Fred Cohen, Barry Hess, Corbin Stewart, and Computer Security.
  7. Per standard CCD procedure, use an anonymizer service when accessing the general Internet.

A.3 Study issues

Eric Thomas and some outside assistants will be doing a set of observations and surveys of the exercises with the goal of understanding how red teams work and develop over time.

- They will observe what you do and write things down.
- They can ask you questions at any time and you should answer honestly.
- You may not ask them questions and he is not allowed to answer them.
- At the end of each exercise fill out the computerized questionnaire at http://10.0.5.53/ (from the gray network). Individual results will only be available to the researchers doing the study and only summary statistics will be published.
- After the form filling out, you will be asked the same questions as a team and will discuss the results as a team to generate additional data.
- Answer questions as honestly as you can.

- We will be recording keystrokes and possibly other information during the study. Please do not subvert or pander to this. This allows us to study technical aspects of the process in detail later.
- Every red team in every exercize may have: (1) Different systems (2) Different protections (3) Different content in the systems (4) Different team objectives

This experimental design is set up to allow repeatable experiments and to allow teams to make staggared starts and stops if necessary. It will also allow us to run the same exercizes on other groups. If you reveal specifics of these exercises, it may invalidate future experiments.

A.4 Operations

In each exercise, there will be access to three networks:

- Internet access will be provided via the Red Network on two systems (the red net)
- CCD access will be provided on two systems (the gray net)
- RedTeam Net access will be provided on two systems. (the green net)

Initially, a standard CCD distribution will be provided for the green net computers and those computers will be attached to a hub that is not yet connected to the green net. At the start of the exercise, it is the job of the team to proceed as they see fit.

Transfer of information from the Internet to the CCD net will function through the Red to Gray Diode (place the files in //graynet/diode on the red net and they will appear in //rednet/diode on the gray net) and transfer from the CCD net to the green net will go through the Gray to Green diode (place the files in //greennet/diode on the gray net and they will appear in //graynet/diode on the green net). No reverse transfers will be allowed. A printer will be available on the green net as well.

Appendix B

Red Team Questionnaire Form

This is the Red Teaming Questionnaire Form that was used during the deception exercises described in Chapter 3. An online version of this form is available at:

`http://all.net/journal/deception/experiments/form.html`

B.1 Questionnaire

Please answer the questions fully and truthfully. Questions with rated answers ranging from 1 to 5 indicate 1 as the least, 5 as the most

1. Enter your name: ________________________________

---

2. Describe the objectives of the exercise:

    ____

    ____

---

3. Did one or more team leaders emerge for your team? Y N
4. If so, who?________________________________
5. Was there a structure to the way your group worked? Y N
6. If so, what was the structure?

    ____

    ____

7. How effective was your teamwork? 1 2 3 4 5
8. How did the group make decisions?

---

9. Did you begin with a strategy or plan for accomplishing your task? Y N
10. If so, What was the strategy?

----

11. How important was this strategy to your success? 1 2 3 4 5
12. How well did the strategy work? 1 2 3 4 5
13. Did any strategy change or emerge while persuing your task? Y N
14. If so, What was the new strategy?

----

15. How important was this strategy to your success? 1 2 3 4 5
16. How well did the strategy work? 1 2 3 4 5
17. Why did you change strategies?

----

18. Did you get stuck anywhere? Y N
19. If so, describe?

---

20. To what extent did you succeed in your task? 1 2 3 4 5
21. How important was it to succeed? 1 2 3 4 5

---

22. What tools did you use? For each tool, provide a number from 1 to 5 indicating the effectiveness of the tool.

----

23. What would you do next if you had more time?

----

24. What would you do differently if you had to do this exercise over again?

----

25. Rate time pressure? `1 2 3 4 5`
26. Rate your uncertainty about how to proceed? `1 2 3 4 5`
27. Was there an exciting moment during the task? `Y N`
28. If so, describe?

----

29. Rate the level of distractions you faced? `1 2 3 4 5`
30. How tiring was this exercise? `1 2 3 4 5`
31. Did you become bored with your task at any time? `Y N`
32. If so, describe?

---

33. Did you identify any defenses? `Y N`
34. If so, what defenses did you identify? For each one, provide a number 1 to 5 indicating how hard they were to identify. Also provide how you identified each one.

----

35. Did you defeat any defenses? `Y N`

36. If so, what defenses did you defeat? For each one, provide a number 1 to 5 indicating how hard they were to defeat. Also provide how you defeated each one.

---

37. How hard was this exercise? `1 2 3 4 5`

38. How interesting was this exercise? `1 2 3 4 5`

39. How enjoyable was this exercise? `1 2 3 4 5`

---

40. What was the least surprising thing you observed?

41. What was the most surprising thing you observed?

42. How surprising was it? `1 2 3 4 5`

---

43. Other comments:

Appendix C

Red Team Data

The data fields in Table C.1 and Table C.2 comprise numerical responses to the following question areas: Date, Deception (Yes or No) Identification, Teamwork effectiveness, Strategy import, Strategy effectiveness, New strategy import, New strategy effectiveness, Extent of success, Import of success, Time pressure, Uncertainty, Distractions, Exhaustion, Difficulty, Interest level, Enjoyability, and Surprise. Detailed questions are included in the "Red Teaming Questionnaire Form" in Appendix B.

Figures C.1, C.2 and C.3 show more detailed information about the statistics obtained from the data.

Date	D	ID	Team	SI	SW	NSI	NSW	Suc	ISuc	Time	Unc	Dist	Tired	Hard	Int	Joy	Surp
2001-06-04	N	JD	3	1	1	3	3	2	4	3	4	1	3	3	3	3	5
2001-06-04	N	JR	3	3	2	3	3	2	2	3	5	3	4	5	1	3	2
2001-06-04	N	SM	3	4	2	4	3	2	5	3	4	3	3	3	3	2	4
2001-06-05	N	JD	5	5	5	3	3	5	5	2	2	2	2	1	1	3	4
2001-06-05	N	OO	5	5	4	4	4	5	5	5	3	2	3	3	4	4	5
2001-06-05	N	MC	5	4	5	3	3	5	5	3	4	2	1	2	4	4	4
2001-06-06	N	MP	4	2	2	3	2	2	3	3	2	2	4	5	4	3	4
2001-06-06	N	CK	3	1	1	3	3	1	3	2	2	4	2	4	3	3	4
2001-06-06	N	JA	3	3	3	3	3	2	3	3	4	2	3	4	4	4	3
2001-06-06	N	GS	2	3	2	3	2	2	5	5	4	2	4	4	5	5	3
2001-06-07	Y	GG	4	3	4	3	3	3	3	2	3	2	2	3	2	2	2
2001-06-07	Y	RW	3	3	4	3	3	5	5	1	4	2	1	2	3	4	3
2001-06-07	Y	SD	4	3	3	5	3	4	5	1	4	2	4	1	2	3	4
2001-06-07	Y	JS	3	4	3	3	3	4	3	2	5	2	3	4	4	3	4
2001-06-08	Y	DH	2	3	3	3	5	4	3	1	4	2	2	3	3	3	5
2001-06-08	Y	AC	2	2	5	3	4	5	4	3	3	3	3	2	2	2	3
2001-06-08	Y	LD	3	2	5	3	3	2	3	5	4	3	2	4	5	4	3
2001-06-08	Y	LA	3	3	3	3	5	5	5	2	4	2	2	3	3	4	3
2001-06-11	Y	JD	2	3	1	3	1	1	5	1	1	1	4	5	3	3	5
2001-06-11	Y	SM	3	3	1	1	1	1	5	3	4	1	4	4	3	2	3
2001-06-11	Y	JR	1	1	2	1	2	3	3	4	1	4	2	4	2	3	2
2001-06-12	N	JD	4	3	3	3	3	3	3	3	4	3	3	5	3	3	3
2001-06-12	N	OO	3	3	2	3	3	2	3	3	3	3	4	4	4	3	3
2001-06-12	N	PS	3	3	3	3	3	3	3	3	3	3	3	5	5	4	4
2001-06-12	N	MC	3	3	2	3	3	2	5	4	5	4	2	5	3	3	4
2001-06-13	N	MP	3	4	4	3	3	3	5	2	3	3	3	5	4	3	3
2001-06-13	N	CK	3	2	1	2	1	1	4	3	4	4	3	4	3	2	3
2001-06-13	N	GS	3	5	2	3	2	1	5	5	3	3	5	5	5	5	3
2001-06-13	N	JA	3	3	2	1	1	2	3	2	3	2	4	4	3	3	3
2001-06-14	Y	GG	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2001-06-14	Y	RW	3	3	2	2	2	2	5	3	4	2	3	4	4	4	3
2001-06-14	Y	JS	4	3	4	3	3	3	3	3	2	4	3	4	4	4	3
2001-06-14	Y	SD	3	4	2	3	3	3	3	2	5	3	2	4	2	1	3
2001-06-15	N	AC	1	1	1	1	1	1	5	5	5	5	5	5	1	2	3
2001-06-15	N	DH	2	1	1	3	3	1	2	4	3	3	3	4	3	3	3
2001-06-15	N	LD	3	2	2	2	2	1	3	1	3	1	4	5	3	3	4
2001-06-15	N	LA	1	1	1	2	1	2	3	3	2	3	2	4	3	2	4
2001-06-15	N	BB	3	1	1	3	2	2	3	1	1	1	3	4	2	2	1
2001-06-15	N	CK	4	4	3	3	3	3	3	3	3	3	3	4	4	3	3
2001-06-18	N	SM	5	5	4	3	3	4	5	4	1	1	4	4	5	5	4
2001-06-18	N	JD	3	3	3	3	3	4	5	5	2	1	1	3	4	4	3
2001-06-18	N	JR	2	2	3	1	4	3	2	3	5	3	3	4	5	1	4
2001-06-19	N	JD	5	4	4	4	4	4	5	3	3	1	1	3	3	3	4
2001-06-19	N	OO	4	4	4	4	3	4	5	3	3	2	3	3	4	3	4
2001-06-19	N	PS	4	4	3	3	3	3	4	4	3	3	1	3	4	4	3
2001-06-19	N	MC	5	4	5	4	5	4	5	5	2	1	1	2	4	4	5
2001-06-20	Y	AN	4	4	4	3	3	3	5	3	3	3	3	5	5	5	3
2001-06-20	Y	GS	4	4	4	3	3	3	5	3	3	3	3	4	5	3	5
2001-06-20	Y	JA	3	3	2	3	3	2	4	3	2	2	3	4	4	4	4
2001-06-20	Y	MP	4	4	2	3	3	2	4	3	3	2	3	5	5	4	4
2001-06-20	Y	CK	4	4	4	3	3	3	4	3	4	2	2	5	4	4	4
2001-06-21	Y	GG	4	2	3	3	3	4	3	3	3	3	3	3	3	5	3
2001-06-21	Y	RW	3	3	3	3	3	2	4	4	3	2	4	4	5	5	3
2001-06-21	Y	JS	4	3	1	3	3	2	3	4	3	1	2	4	5	4	3
2001-06-21	Y	SD	3	4	2	3	3	2	4	2	5	4	4	4	3	2	2
2001-06-21	Y	VN	4	3	3	3	3	2	3	2	4	1	2	5	5	5	3
2001-06-22	N	AC	3	1	3	3	3	1	5	4	2	5	3	5	3	3	3
2001-06-22	N	LD	1	1	1	1	1	2	3	2	1	5	5	5	4	3	3
2001-06-22	N	DH	3	2	1	5	1	2	1	2	2	5	2	4	5	3	4
2001-06-22	N	BB	3	3	3	3	3	3	3	2	3	1	3	3	4	4	3
2001-06-22	N	CK	3	1	1	1	1	2	4	3	5	5	5	4	4	2	3
2001-06-22	N	LA	2	2	1	3	3	1	1	1	2	2	3	4	3	2	4

Table C.1: Data on Confounding Factors (part 1)

Date	D	ID	Tm#	SI	SW	NSI	NSW	Suc	ISuc	Time	Unc	Dist	Tired	Hard	Int	Joy	Surp
2001-06-25	Y	SM	4	4	4	3	3	3	5	4	2	3	3	4	4	3	1
2001-06-25	Y	SD	4	3	2	3	3	2	3	3	3	2	2	4	4	4	3
2001-06-25	Y	BB	2	2	1	3	3	1	3	2	4	1	3	5	3	2	3
2001-06-25	Y	JD	3	1	1	1	1	1	3	3	3	1	3	5	3	2	3
2001-06-25	Y	JR	3	3	3	2	3	2	2	1	3	1	2	4	3	3	3
2001-06-25	Y	KM	3	1	1	1	1	1	5	5	5	2	2	5	5	5	3
2001-06-26	N	MP	3	4	3	3	3	3	3	3	4	2	3	4	4	3	4
2001-06-26	N	GS	2	2	2	3	3	2	4	4	4	2	4	5	3	3	3
2001-06-26	N	PS	3	3	3	3	3	3	4	4	4	4	3	4	3	3	3
2001-06-26	N	JA	3	4	4	3	3	4	4	4	3	2	2	4	5	4	4
2001-06-26	N	NP	4	4	2	4	4	2	3	4	3	3	5	5	4	3	2
2001-06-26	N	MC	5	5	4	3	3	4	5	4	4	4	1	4	3	4	4
2001-06-27	N	GG	4	3	3	4	3	4	3	3	3	3	3	3	3	3	3
2001-06-27	N	RW	3	5	2	3	2	2	4	3	3	2	1	5	5	3	4
2001-06-27	N	JS	5	4	3	4	3	4	3	2	3	2	1	3	5	5	3
2001-06-27	N	AN	4	3	3	3	3	3	5	3	4	5	3	5	4	4	3
2001-06-27	N	VN	5	5	5	3	3	3	3	1	4	3	3	4	3	3	2
2001-06-27	N	OO	3	5	3	4	3	2	5	4	3	3	3	3	3	3	3
2001-06-28	Y	CK	3	2	2	3	3	2	4	3	4	4	3	4	3	4	4
2001-06-28	Y	RY	4	1	1	1	1	2	2	2	4	3	3	4	5	3	4
2001-06-28	Y	BS	2	4	3	3	3	2	4	3	3	3	2	4	5	3	3
2001-06-28	Y	NB	3	3	3	3	2	3	3	3	3	3	2	4	5	4	3
2001-06-29	Y	JD	4	3	3	3	3	3	3	4	4	3	3	5	1	1	3
2001-06-29	Y	DH	3	3	3	3	3	3	3	1	3	2	3	4	3	3	3
2001-06-29	Y	LA	3	2	1	3	2	1	4	3	3	2	2	5	3	2	4
2001-06-29	Y	CK	3	3	3	2	3	3	4	2	4	3	3	3	5	3	5
2001-06-29	Y	LD	2	1	1	3	1	2	2	1	5	2	5	3	2	1	3
2001-07-09	Y	JD	2	1	1	1	1	1	5	1	4	3	3	5	3	2	3
2001-07-09	Y	JR	3	3	3	3	3	1	2	1	5	5	5	5	1	1	4
2001-07-09	Y	SM	3	3	2	3	3	3	3	3	3	5	5	5	3	1	3
2001-07-09	Y	JC	3	5	5	5	2	3	5	1	4	1	4	4	4	4	5
2001-07-09	Y	BB	2	2	1	1	1	1	3	2	5	4	3	5	3	2	3
2001-07-09	Y	KM	2	2	1	1	1	1	3	1	4	4	5	4	1	2	4
2001-07-09	Y	SD	4	4	3	3	2	2	4	2	2	4	3	4	4	4	3
2001-07-10	N	NP	4	3	3	3	3	4	5	4	3	2	3	5	4	4	4
2001-07-10	N	JA	4	3	3	3	3	3	4	4	3	2	3	4	4	4	4
2001-07-10	N	GS	5	3	3	3	3	5	5	5	3	3	5	5	5	3	3
2001-07-10	N	PS	3	3	3	3	3	2	4	3	4	3	3	4	3	2	3
2001-07-10	N	MP	4	4	3	3	3	3	5	4	4	2	3	5	4	3	3
2001-07-10	N	MC	2	2	1	3	3	1	4	2	5	3	4	4	2	1	3
2001-07-11	N	GG	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2001-07-11	N	JS	2	3	2	3	2	2	3	2	4	3	3	5	2	2	2
2001-07-11	N	LA	3	3	3	3	3	2	3	3	1	2	4	5	2	1	1
2001-07-11	N	VN	3	3	3	3	3	1	3	1	5	3	3	5	2	2	3
2001-07-11	N	OO	4	4	3	1	1	3	4	2	3	3	3	4	3	3	3
2001-07-12	Y	NB	4	3	1	3	1	3	3	1	2	3	3	5	4	5	4
2001-07-12	Y	RY	3	3	2	3	3	1	4	2	4	4	3	5	3	3	3
2001-07-12	Y	CK	2	4	1	3	3	1	4	2	4	5	2	4	3	4	3
2001-07-12	Y	BS	4	3	1	3	3	1	3	3	3	2	3	5	4	3	3
2001-07-16	Y	BB	3	3	3	3	3	2	2	3	3	1	3	4	2	2	3
2001-07-16	Y	JD	5	5	5	3	3	3	3	1	3	1	1	5	5	5	3
2001-07-16	Y	SD	4	4	4	3	3	4	4	2	4	4	3	4	3	2	3
2001-07-17	N	NP	3	4	3	3	3	3	4	4	4	3	5	4	1	1	4
2001-07-17	N	JA	3	3	2	3	3	3	2	2	3	2	4	4	2	2	3
2001-07-17	N	PS	5	4	4	3	3	3	4	4	3	3	1	3	4	4	4
2001-07-17	N	MC	3	4	2	3	3	2	5	4	3	3	1	4	4	3	4
2001-07-17	N	GS	5	3	3	5	3	3	5	5	3	3	3	3	5	3	5
2001-07-18	N	VN	4	3	3	3	3	3	3	2	4	1	3	5	3	3	3
2001-07-18	N	AN	4	3	3	3	3	3	4	3	3	1	3	5	4	3	3
2001-07-18	N	OO	4	3	3	3	2	3	4	1	3	1	2	4	3	3	3

Table C.2: Data on Confounding Factors (part 2)

All Actors Hour 1		All Actors Hours 2-4	
Duration (sec)	Freq	Duration (sec)	Freq
0	0	0	0
2	14	2	18
4	30	4	8
6	27	6	9
8	32	8	7
10	10	10	36
12	14	12	14
14	15	14	10
16	15	16	6
18	5	18	8
20	8	20	37
22	4	22	4
24	6	24	2
26	4	26	6
28	3	28	2
30	5	30	24
32	4	32	3
34	1	34	3
36	2	36	6
38	2	38	1
40	2	40	27
42	2	42	1
44	1	44	2
50	1	46	2
		48	3
		50	11
		52	2
		58	2
		60	3
		66	1
		70	3
		80	1
		84	1
		92	1
		120	1

Figure C.1: Statistics – Durations

Summary Statistics All Actors Hour 1

Column1		Column2	
Mean	23.16666667	Mean	8.5
Standard Error	2.948634335	Standard E	1.937427
Median	23	Median	4.5
Mode	#N/A	Mode	1
Standard Deviat	14.44529912	Standard D	9.491415
Sample Varianc	208.6666667	Sample Va	90.08696
Kurtosis	-1.06098868	Kurtosis	1.214585
Skewness	0.080505419	Skewness	1.457951
Range	50	Range	32
Minimum	0	Minimum	0
Maximum	50	Maximum	32
Sum	556	Sum	204
Count	24	Count	24

Summary Statistics All Actors Hours 2-4

Column1		Column2	
Mean	38.11428571	Mean	7.4
Standard Error	4.742785336	Standard Erro	1.606810225
Median	34	Median	3
Mode	#N/A	Mode	1
Standard Deviation	28.05852228	Standard Devi	9.506035235
Sample Variance	787.2806723	Sample Varian	90.36470588
Kurtosis	0.81103039	Kurtosis	3.8559586
Skewness	0.94185868	Skewness	2.114464743
Range	120	Range	37
Minimum	0	Minimum	0
Maximum	120	Maximum	37
Sum	1334	Sum	259
Count	35	Count	35

Figure C.2: Statistics – Summary

t-Test: Two-Sample Assuming Unequal Variances
Duration of Response All Actors

	Hour 1	Hours 2-4
Mean	23.1666667	38.11429
Variance	208.666667	787.2807
Observations	24	35
Hypothesized Mean Difference	0	
df	54	
t Stat	-2.67656028	
P(T<=t) one-tail	0.0049144	
t Critical one-tail	1.67356575	
P(T<=t) two-tail	0.00982881	
t Critical two-tail	2.00488103	

Figure C.3: Statistics – t-Test

Distribution

1	MS 0151	T. Hunter, 9000
1	MS 0151	A. Ratzel, 9750
1	MS 9001	J. Handrock, 8000
1	MS 9003	C. Hartwig, 8940
1	MS 9003	K. Washington, 8900
1	MS 9011	N. Durgin, 8941
1	MS 9011	B. Hess, 8941
1	MS 9011	J. Howard, 8941
1	MS 9913	D. Rogers, 8941
3	MS 9018	Central Technical Files, 8945-1
1	MS 0899	Technical Library, 9616
1	MS 9021	Classification Office, 8511/Technical Library, NS 0899, 9616
		DOE/OSTI via URL
1	MS 0323	D.L. Chavez, LDRD Office, 1011